

Dissertation zur Erlangung des Doktorgrades  
der Fakultät für Chemie und Pharmazie  
der Ludwig-Maximilians-Universität München

**Chemical Synthesis and Enzymatic  
Incorporation of Artificial Nucleotides**

**Chemische Synthese und enzymatischer  
Einbau von künstlichen Nukleotiden**

Sascha Serdjukow  
geboren in Haldensleben, Deutschland

München, 2016

## **Erklärung**

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Thomas Carell betreut.

## **Eidesstattliche Versicherung**

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, .....

.....  
Sascha Serdjukow

Dissertation eingereicht: 18.01.2016

1. Gutachter: Prof. Dr. Thomas Carell

2. Gutachter: Prof. Dr. Anja Hoffmann-Röder

Mündliche Prüfung: 18.04.2016

*„Sei naiv und mach´ ein Experiment.“*

Feodor Lynen

*„Auch eine Enttäuschung, wenn sie nur gründlich und endgültig ist,  
bedeutet einen Schritt vorwärts.“*

Max Planck

## Publications

M. Tomás-Gamasa, S. Serdjukow, M. Su, M. Müller, T. Carell, *Angew. Chem. Int. Ed.* **2015**, 54, 796-800. *"Post-It" Type Connected DNA Created with a Reversible Covalent Cross-Link.*

S. Serdjukow, F. Kink, B. Steigenberger, M. Tomás-Gamasa, T. Carell, *Chem. Commun.*, **2014**, 50, 1861-1863. *Synthesis of  $\gamma$ -labeled nucleoside 5'-triphosphates using click chemistry.*

M. Su, M. Tomás-Gamasa, S. Serdjukow, P. Mayer, T. Carell, *Chem. Commun.* **2014**, 50, 409-411. *Synthesis and properties of a Cu(II) complexing pyrazole ligand in DNA.*

## Further Publications

J. Gajewski, F. Buelens, S. Serdjukow, N. Cortina, H. Grubmueller, M. Grninger, *Nat. Chem. Bio.*, manuscript under revision. *Engineering fatty acid synthases (FAS) for directed polyketide production.*

H. Staudt, M.G. Hoesl, A. Dreuw, S. Serdjukow, D. Oesterhelt, N. Budisa, J. Wachtveitl, M. Grninger, *Angew. Chem., Int. Ed.* **2013**, 32, 8463-8466. *Directed manipulation of a flavoprotein photocycle.*

## Conference Presentations

Poster Presentation: "Enzymatic Incorporation of a Reversible Covalent DNA Cross-link", CAS Conference Synthetic Biology, Munich, Germany, July 2015.

Poster Presentation: "Enzymatic Incorporation of a Reversible Covalent DNA Cross-link", VII. Nukleinsäurechemie-Treffen, Berlin, Germany, September 2015.

## Danksagungen

Mein erster Dank gilt meinem Betreuer Prof. Dr. Thomas Carell für das Gewähren immenser akademischer Freiheit, wodurch die Neugier, zum Antrieb aller Bestrebungen werden konnte. Das damit verbundene Vertrauen an mich, diese Freiheit sinnvoll und gewissenhaft zu nutzen, habe ich immer als Wertschätzung empfunden. Für mich persönlich war die enorme Unterstützung und das Verständnis für die Schwierigkeiten, die mit einer Familiengründung verbunden sind, von unschätzbarem Wert.

Bei Frau Prof. Dr. Anja Hoffmann-Röder bedanke ich mich für die Zweitkorrektur der Dissertation und den Mitgliedern der Prüfungskommission bin ich für die Begutachtung der Arbeit und der Teilnahme am Rigorosum zu Dank verpflichtet.

Den deutschen Steuerzahlern möchte ich für die Finanzierung von akademischer Forschung danken. Die finanziellen Mittel tragen nicht nur zur Ausbildung von Wissenschaftlern bei, sondern leisten auch einen wichtigen Beitrag zur technologischen Weiterentwicklung der Gesellschaft und letztendlich der Menschheit. Die Gelegenheit durch ein Forschungsprojekt, einen winzigen Baustein zum Wissen der Menschheit hinzufügen zu dürfen, muss global betrachtet angesichts existenzieller Krisen als besonderes Privileg angesehen werden.

Herrn Dr. Markus Müller danke ich für die Hilfe bei den Kristallisationsexperimenten und bei der Überwindung alltäglicher kleiner Hindernisse, die im Laboralltag auftreten.

Für die gute Zusammenarbeit beim Projekt der künstlichen Basenpaare danke ich Dr. Maria Tomás-Gamasa und Meng Su.

Den Kollegen aus dem Labor F4.017 verdanke ich abwechslungsreiche Musik, Hilfsbereitschaft und Diskussionen von wissenschaftlichen Fragestellungen. Dies und auch Unternehmungen außerhalb der Arbeit haben zu Verbundenheit und Freundschaft geführt. Den Mitgliedern der Arbeitsgruppe gebührt besonderer Dank für ihre Bereitschaft zum Teilen von Wissen und der Vermittlung von Fähigkeiten.

Auch meinen Studenten Matthias K., Alexander N., Kerstin S., Lorenz K., Tobias B. und Patricia L. bin ich für ihr Engagement und ihre Neugier bei der Durchführung von Experimenten in Rahmen von Forschungspraktika und Bachelorarbeiten zu Dank verpflichtet.

Für sorgfältiges Korrekturlesen und kritische Kommentare bin ich Edris Parsa, Ilka Sührer, Barbara Steigenberger, Thomas Wildenhof, Arne Schröder, Korbinian Brunner, Florian Kink, Dr. Iacovos Michaelides und Dr. Jakob Franke zu Dank verpflichtet.

Meiner Freundin danke ich für die gemeinsame Zeit mit zahlreichen herrlichen Wanderungen, der Geburt unseres Sohnes und vielen weiteren Dingen. Die Unterstützung und das gerechte Aufteilen von Arbeit und Verpflichtungen haben es uns ermöglicht, unser Kind, die jeweilige Promotion und eine erfüllte Partnerschaft unter einen Hut zu bringen.

Meinem Sohn Martin möchte ich für sein fröhliches und lebhaftes Gemüt danken. Alle Anstrengungen und Entbehrungen kann er mit seinem Lächeln und seiner unendlichen Neugier wegwischen.

Meinen Eltern bin ich dankbar für Ihren Glauben in meine Fähigkeiten und die Unterstützung während des Studiums, ohne die Vieles schwieriger bis unmöglich geworden wäre.

## Table of Contents

<b>Summary .....</b>	<b>IV</b>
<b>1 Introduction .....</b>	<b>1</b>
1.1 DNA .....	1
1.2 Nucleoside Chemistry .....	2
1.3 Chemical (Tri)phosphorylation Methods .....	4
1.4 Cu(I)-Catalyzed Alkyne-Azide Cycloaddition .....	9
1.5 Application of Artificial Nucleosides and Nucleotides .....	13
1.5.1 Therapeutic Nucleosides and Nucleotides .....	13
1.5.2 Unnatural Base Pairs – Alternative Hydrogen Bonding .....	14
1.5.3 Hydrophobic Unnatural Base Pairs .....	16
1.5.4 Metal-Base Pairs .....	19
<b>2 Aim of the Project .....</b>	<b>22</b>
<b>3 Results and Discussion .....</b>	<b>24</b>
<b>Part I – Efforts Towards Faithful Transcription and Translation of a Metal Base Pair .....</b>	<b>24</b>
3.1 Synthesis of a Salicylaldehyde Ribonucleotide .....	24
3.2 Synthesis of a Benzaldehyde Ribonucleotide .....	28
3.3 Synthesis of a Salicylaldehyde Ribophosphoramidite .....	32
3.4 Transcription Experiments with STP and AldTP .....	38
3.5 Transcription Experiments using T7 RNA Polymerase Mutants .....	43
<b>Part II – A Covalent Base Pair .....</b>	<b>47</b>
3.6 Overview of the Synthesized Amine and Aldehyde Phosphoramidites .....	47
3.7 Melting Temperature Analysis of Strands Containing Amine and Aldehyde Bases .....	49
3.8 Synthesis of a Salicylaldehyde and an Aromatic Amine Deoxyribonucleotide .....	55
3.9 Primer Extension Experiments with dSTP and dT <sub>0</sub> TP .....	59
3.9.1 Primer Extensions with dSTP .....	59
3.9.2 Primer Extensions with dT <sub>0</sub> TP .....	62
3.10 Co-crystallization of <i>Bst</i> Pol I with S:T <sub>0</sub> Containing DNA .....	64
3.11 Chemical Efforts to Improve the Enzymatic Incorporation Efficiency .....	68
3.11.1 Synthesis of a Naphthalene Amine Base Triphosphate .....	68
3.11.2 Primer Extension Experiments with dNaaTP .....	71
3.11.3 Design and Proposed Synthesis of Further Improved Covalent Base Pairs .....	77
<b>Part III – Click Chemistry Labeling of Phosphate-modified Nucleotides .....</b>	<b>79</b>
3.12 Synthesis of $\gamma$ -Labeled Nucleotides .....	79
3.13 Enzymatic Incorporation of $\gamma$ -Labeled Nucleotides into DNA .....	85

<b>3.14</b>	<b>Enzymatic Incorporation of <math>\gamma</math>-Labeled Nucleotides into RNA .....</b>	<b>86</b>
<b>3.15</b>	<b>Enzymatic Labeling Efforts Involving <math>\gamma</math>-Labeled Nucleotides .....</b>	<b>90</b>
<b>3.16</b>	<b>Synthesis and Application of an <math>\alpha</math>-Alkyne Deoxyguanosine 5'-Triphosphate.....</b>	<b>91</b>
<b>4</b>	<b>Conclusions and Outlook.....</b>	<b>95</b>
<b>5</b>	<b>Experimental Part.....</b>	<b>98</b>
<b>5.1</b>	<b>General Methods and Materials for Synthesis.....</b>	<b>98</b>
<b>5.2</b>	<b>Chemical Synthesis.....</b>	<b>100</b>
5.2.1	Preparation of the Triphosphate Reagent Solution.....	100
5.2.2	Synthesis of the Salicylaldehyde Ribonucleotide.....	101
5.2.3	Synthesis of the Benzaldehyde Ribonucleotide.....	107
5.2.4	Synthesis of a Salicylaldehyde Ribophosphoramidite.....	112
5.2.5	Synthesis of the Salicylaldehyde Nucleoside and Phosphoramidite .....	119
5.2.6	Synthesis of dSTP.....	119
5.2.7	Synthesis of dT <sub>0</sub> TP.....	120
5.2.8	Synthesis of a Naphthalene Amine Triphosphate.....	123
5.2.9	Synthesis of $\gamma$ -Alkyne Labeled Nucleotides.....	130
5.2.10	Synthesis of Fluorophore Azides.....	137
5.2.11	Synthesis of $\gamma$ -Fluorophore Labeled Nucleoside Triphosphates .....	140
5.2.12	Synthesis of an $\alpha$ -Alkyne Labeled Nucleoside Triphosphate .....	143
<b>5.3</b>	<b>Oligonucleotide Synthesis .....</b>	<b>145</b>
<b>5.4</b>	<b>Biochemical Experiments .....</b>	<b>147</b>
5.4.1	Buffers .....	147
5.4.2	Chemically Competent <i>E. coli</i> Strains.....	148
5.4.3	DNA Oligonucleotide Sequences.....	148
5.4.4	Melting Curve Experiments.....	150
5.4.5	Site-Directed Mutagenesis.....	150
5.4.6	Transformation .....	151
5.4.7	Plasmid DNA Preparation .....	151
5.4.8	DNA sequencing .....	151
5.4.9	Determination of Protein and DNA Concentrations.....	151
5.4.10	Gel Electrophoresis of Proteins, DNA and RNA .....	152
5.4.11	Expression and Purification of T7 RNA Polymerase Mutants .....	152
5.4.12	Transcription Assay Conditions .....	153
5.4.13	Expression and Purification of <i>Bst</i> Pol I.....	154
5.4.14	DNA- <i>Bst</i> Pol I Co-Crystallization.....	154
5.4.15	Primer Extension Experiments .....	157



---

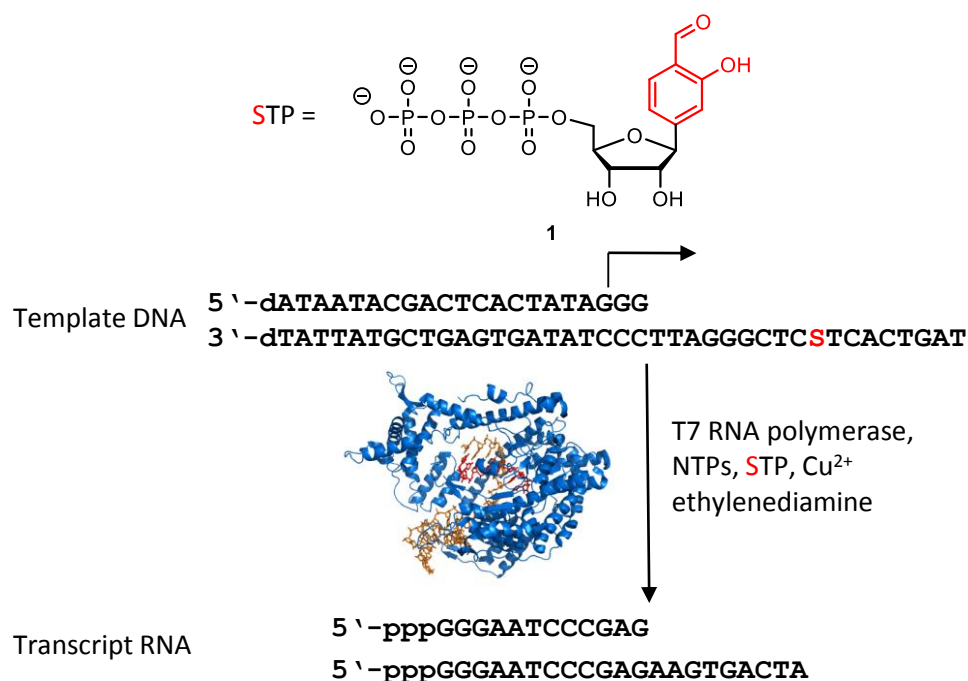
<b>6</b>	<b>Abbreviations .....</b>	<b>158</b>
<b>7</b>	<b>Appendix .....</b>	<b>162</b>
<b>7.1</b>	<b>Protein Sequences .....</b>	<b>162</b>
<b>7.2</b>	<b>DNA Sequences.....</b>	<b>162</b>
<b>8</b>	<b>References .....</b>	<b>163</b>

## Summary

Artificial nucleosides and nucleotides are used in therapy,<sup>[1-2]</sup> biotechnology (e.g. DNA sequencing)<sup>[3]</sup> and fundamental research.<sup>[4]</sup> Compared to their natural congeners, these compounds differ at the nucleobase,<sup>[5]</sup> the sugar<sup>[6]</sup> or the phosphate moiety.<sup>[7]</sup> In course of this work, several artificial nucleotides with nucleobase or phosphate modifications were synthesized and studied for their ability to be incorporated by enzymes.

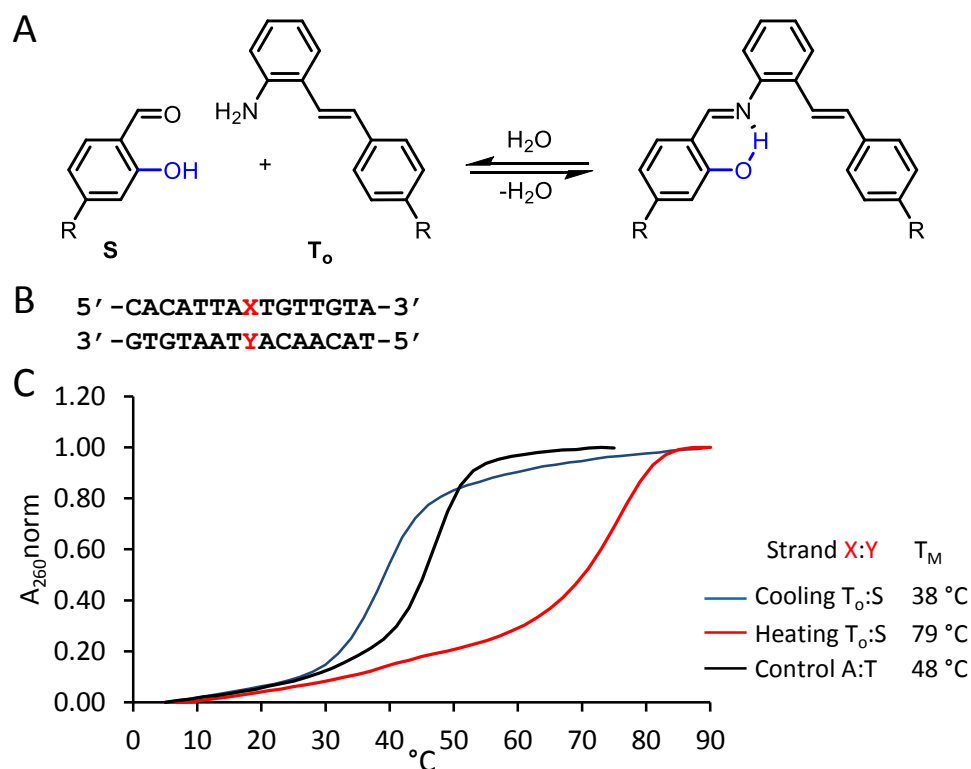
The expansion of the genetic code, to allow coding of additional non-canonical amino acids and thus the creation of novel biocatalysts, is one major goal in synthetic biology. One strategy aims at the creation of an additional artificial base pair, which can be replicated, transcribed and translated by the existing enzyme machinery. In order to achieve this extremely challenging goal, the design of the novel base pair candidate is crucial. Most approaches rely on shape complementarity and  $\pi$ -stacking,<sup>[8-10]</sup> while hydrogen bonds are not necessary for selective interstrand interaction.<sup>[11]</sup> A completely different strategy, the interaction *via* reversible covalent bonding and metal complexation had been developed by *Clever et al.*<sup>[12]</sup> Based on the salen catalyst, a DNA metal base pair had been prepared, which was dependent on  $\text{Cu}^{2+}$  and could be amplified in the polymerase chain reaction.<sup>[13]</sup>

In this research project efforts towards transcription and translation of the salen base pair were analyzed for which several compounds were required. The phosphoramidite of the salicylaldehyde deoxynucleoside was synthesized. A synthesis route to the salicylaldehyde ribonucleoside was established and its triphosphate **1** was prepared successfully. In order to study the role of the *ortho*-hydroxyl group for the imine formation, a benzaldehyde nucleoside 5'-triphosphate (AldTP) and its phosphoramidite were synthesized. Transcription of the salicylaldehyde base pair by T7 RNA polymerase was analyzed, but resulted in transcription termination prior to incorporation of the artificial base or mutation of the artificial salicylaldehyde to an adenine. Neither T7 RNA polymerase mutants nor solid-phase synthesis using a salicylaldehyde ribophosphoramidite could provide RNA containing the salicylaldehyde.



**Scheme 0-1.** Transcription of the salicylaldehyde nucleoside by T7 RNA polymerase results in transcription stop or mutation to a natural adenosine.

Since transcription of the salicylaldehyde metal-base pair was not possible, redesign of the artificial base pair was elaborated. The aim was to develop a base pair, which solely relied on reversible covalent imine chemistry without the need for additional metal chelation. The existing salicylaldehyde was used as the starting point and it provided the aldehyde subunit. A set of different amine nucleoside phosphoramidites was designed and synthesized. Amine nucleoside and aldehyde nucleoside containing DNA strands were produced and employed in melting temperature experiments. The combination of an aromatic amine (T<sub>o</sub>) with the salicylaldehyde (S) provided a melting temperature increase of up to 31 K compared to the control strand with a central adenine:thymine (A:T) pair (Figure 0-1, C). Notably, the duplex containing this aldehyde-aromatic base combination displayed a strong hysteresis effect. Comparison between salicylaldehyde and benzaldehyde containing DNA strands proved that the hydroxyl group directly adjacent to the aldehyde is crucial for imine stability.



**Figure 0-1.** (A) Reversible chemistry enables salicylaldehyde: aromatic amine (S:T<sub>0</sub>) base pair formation. (B) DNA strands containing different base pairs (X:Y) were studied in melting temperature analysis (C).

Enzymatic incorporation of the novel, unnatural base pair candidate by DNA polymerases was studied. Therefore, the salicylaldehyde (dSTP) and aromatic amine triphosphate (dT<sub>0</sub>TP) were synthesized. DNA templates containing the aromatic amine or the salicylaldehyde nucleoside in the template strand were prepared and different DNA polymerases were studied for their ability to incorporate the artificial nucleotides. Using a polymerase mixture, dSTP incorporation opposite the aromatic amine and full elongation of the primer was achieved. Klenow fragment polymerase was able to synthesize the base pair in the reverse case, but subsequent elongation was inefficient after dT<sub>0</sub>TP addition.

In order to gain insight into the structure of the base pair, S:T<sub>0</sub> containing strands were co-crystallized in complex with the DNA polymerase I from *Geobacillus stearothermophilus*. Well-diffracting crystals were obtained within several days and the X-ray structure was solved. The S:T<sub>0</sub> base pair was found completely isosteric in comparison to a natural base pair.

Based on the structure, the aromatic amine was changed to a naphthalene amine. A naphthalene amine triphosphate (dNaaTP) was synthesized and used in primer extension experiments. Single nucleotide incorporations using various polymerases revealed that the naphthalene nucleotide is accepted more readily by the enzymes. Even full elongation of the primer was achieved in the presence of five dNTPs and dNaaTP was incorporated opposite the templating salicylaldehyde. In summary, a reversible covalent base pair was developed which site-specifically crosslinks DNA and thereby stabilizes the

duplex. The base pair can be formed by DNA polymerases and does not disturb the structure of the DNA double strand.

Terminally phosphate modified nucleotides are important substrates in next-generation sequencing<sup>[3]</sup> and real-time methods.<sup>[14]</sup> Despite the increasing importance of these molecules, their synthesis is challenging.<sup>[7]</sup>

A short and efficient synthesis of  $\gamma$ -modified triphosphates was developed based on the Cu(I)-catalyzed alkyne azide cycloaddition (CuAAC). Starting from nucleotides, alkyne labels were selectively introduced on the  $\gamma$ -phosphate with 70-86% yield for all eight major nucleoside triphosphates (Figure 0-2). Subsequent CuAAC with different fluorescent dye azides allowed access to  $\gamma$ -fluorophore labeled nucleotides with 70-77% isolated yield.



**Figure 0-2.** Click chemistry based synthesis of  $\gamma$ -modified nucleoside triphosphates.

The modified deoxynucleotides were accepted by DNA polymerases in primer extension experiments. T7 RNA polymerase mediated labeling of RNA transcripts was achieved *in vitro*. By supplementation of natural GTP with  $\gamma$ -alkyne and  $\gamma$ -fluorescein GTP, 5'-labeling of the RNA transcript was possible. Therefore,  $\gamma$ -labeled nucleotides synthesized by this new method are potential substrates for real-time enzymatic studies and labeling. Efforts to expand the enzymatic labeling strategy and the application of  $\alpha$ -modified nucleotides were yet unsuccessful.

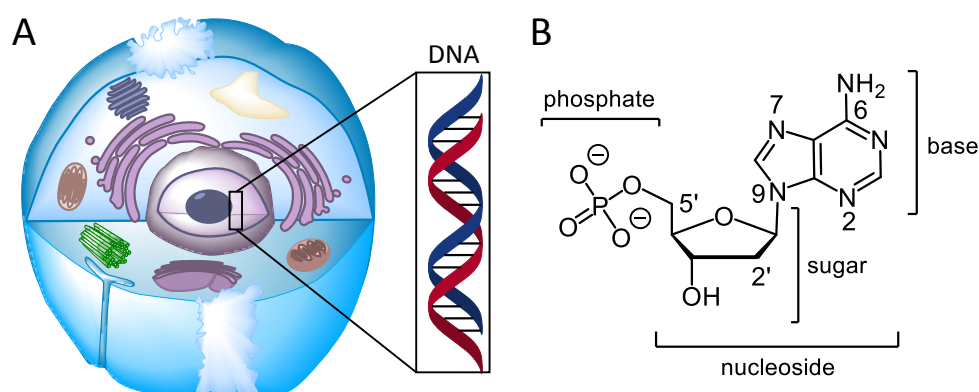


## 1 Introduction

### 1.1 DNA

Despite the biological diversity of organisms, life on the molecular level is surprisingly uniform.<sup>[15]</sup> Many of the biomolecules are biopolymers, which consist of a set of relatively few monomer building blocks. Proteins consist of 20 amino acids, lipids consist of fatty acids and DNA consists of four major nucleotides. DNA is the central molecule of life and the genetic code is universal for all organisms.<sup>[15]</sup>

In 1869, *Friedrich Miescher* isolated a phosphorus rich substance from leukocyte nuclei, which he called “nuclein”.<sup>[16]</sup> This term has been preserved in today's name for the molecule: deoxyribonucleic acid, DNA (Figure 1-1, A). More than 50 years were needed since the initial isolation until the components of DNA were isolated and identified by *Levene*,<sup>[17]</sup> *Klein* and *Thannhauser*.<sup>[18]</sup> Unambiguous structural elucidation of the monomers was strongly aided by chemical synthesis in the group of *Todd*.<sup>[19-20]</sup> DNA is a polymer of four different monomers comprising the bases adenine (A), cytosine (C), guanine (G) and thymine (T). Each of these bases is connected to deoxyribose *via* an hemiaminal to form a nucleoside (Figure 1-1, B). Phosphate diester bonds join the monomers (nucleotides) at the 5'- and 3'-hydroxylgroup of the sugar.<sup>[19]</sup>

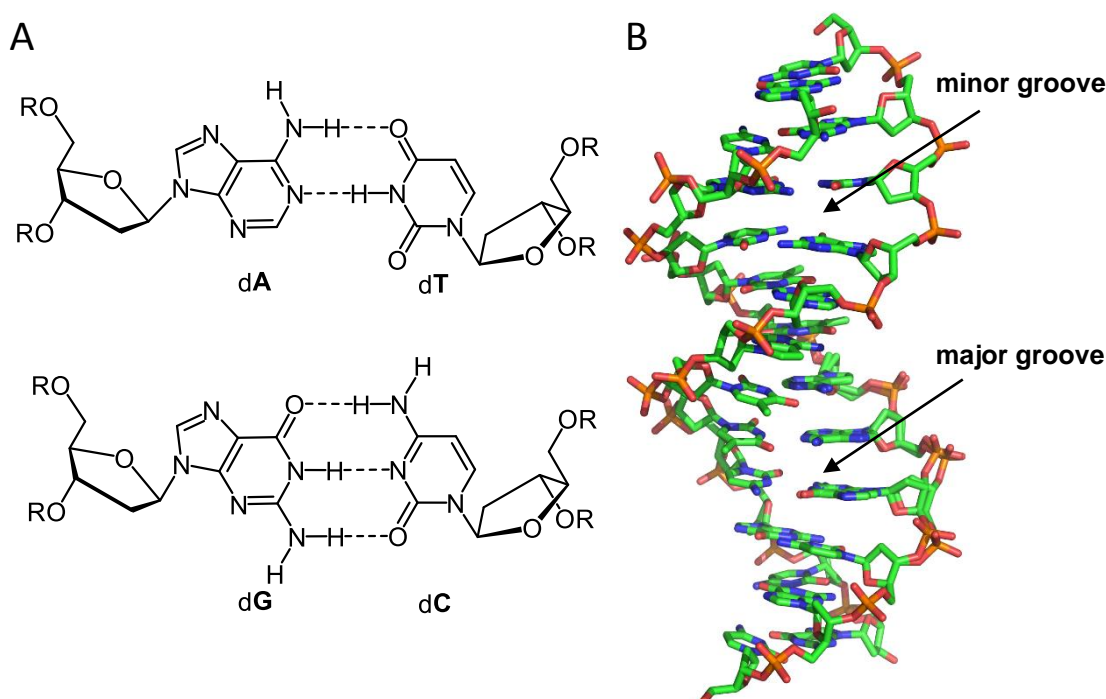


**Figure 1-1.** (A) DNA is located in the nucleus of eucaryotes. (B) Exemplary DNA monomer: 2'-deoxyadenosine 5'-monophosphate.

A major breakthrough was the identification of DNA as the carrier of genetic information by *Avery et al.* in 1944.<sup>[21]</sup> A few years later (1953), the structure of the DNA was deciphered.<sup>[22-24]</sup> A helical structure is adopted from two pairing strands. A sugar-phosphate backbone connects the bases, which selectively base pair *via* hydrogen bonds: A pairs with T and C pairs with G.

DNA is the prime example of the structure-function relation, which is found in many biomacromolecules.<sup>[15]</sup> The order of bases in the DNA strand (primary structure) encodes the genetic information and the reversibility of the base pairs allows replication.<sup>[24-25]</sup> A “transient copy” of this information is generated by transcription of DNA into messenger RNA (mRNA), which is needed for protein synthesis. Three base pairs, the so-called triplet, encodes one amino acid (or stop codon).<sup>[26-27]</sup>

Genomic DNA is mostly found in the right-handed B-conformation and has ten base pairs per turn.<sup>[28]</sup> The major and the minor groove are two areas of the DNA helix that are exposed to the solvent and allow interaction of base pairs with proteins for selective sequence recognition.<sup>[29]</sup>



**Figure 1-2.** (A) Hydrogen bonds mediate the base pairing in double stranded DNA. Two base pairs are formed selectively between A:T and G:C. R = oligonucleotide (B) B-DNA structure, generated from pdb entry: 3OPI.<sup>[30]</sup>

## 1.2 Nucleoside Chemistry

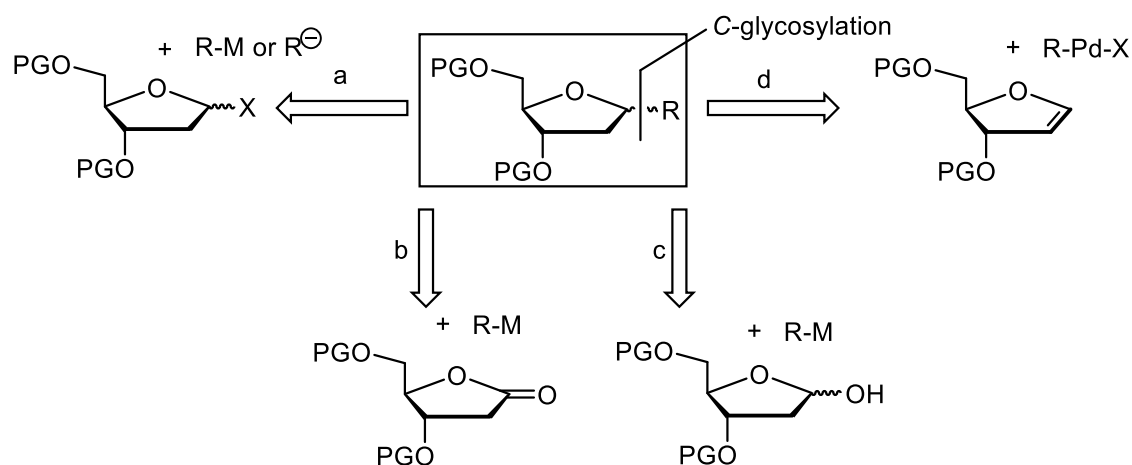
Nucleoside chemistry is the key to study DNA, RNA, nucleosides and nucleotides. It enables the targeted preparation of nucleosides and oligonucleotides to investigate their physicochemical properties and their role in biological systems. Before chemical synthesis of nucleosides was possible, progress depended on isolation of molecules from natural sources and subsequent chemical or enzymatic decomposition.<sup>[26-27]</sup>

Protecting group strategies are an essential part in nucleoside chemistry.<sup>[31-32]</sup> Since the nucleosides contain a variety of functional groups in close proximity, chemo-, regio- and stereoselectivity are serious issues. Protection of functional groups, to prevent side reactions is a common method to obtain the desired selectivity.<sup>[31, 33]</sup>

Two major retrosynthetic strategies are applied: the glycosidic bond formation or construction of the nucleobase on 1'-substituted sugars. The first one being the key step in most nucleoside syntheses. Efficient reaction to a single anomer ( $\alpha$  or  $\beta$ ) and regioselective reaction with the sugar derivative is highly desirable. Several efficient *N*-glycosylation methods have been established, which have been optimized to provide mostly the natural  $\beta$ -anomers at C1'.<sup>[34-35]</sup>

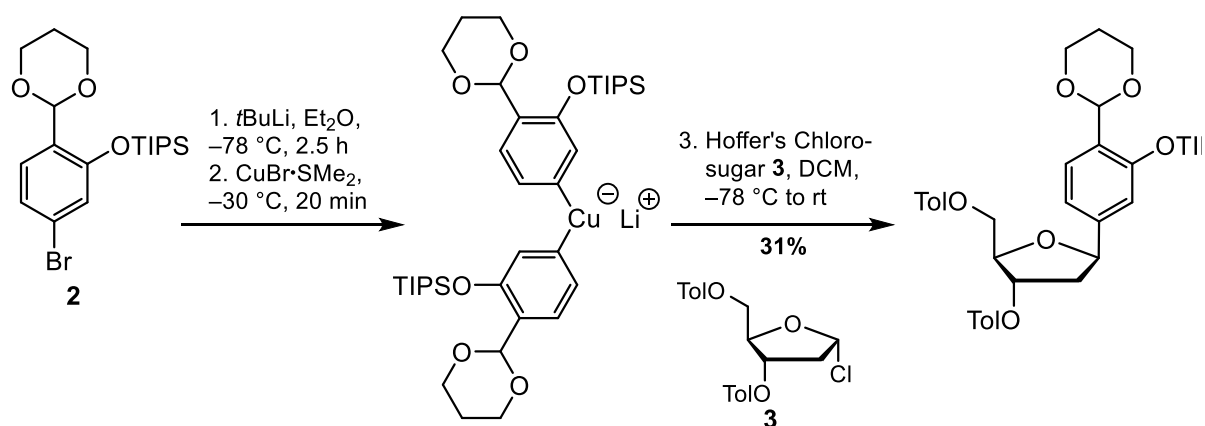


The synthesis of *C*-nucleosides was crucial for this project, therefore some synthetic methods for the preparation of these nucleosides are presented briefly. A typical synthesis strategy for *C*-nucleosides is attack of a nucleophilic base on a sugar electrophile (Scheme 1-1, a-c). Depending on the prefunctionalization of base and sugar, several *C*-glycosylation methods can be distinguished. Halosugars are reacted with organometallic nucleobase derivatives or deprotonated nucleobase precursors (Scheme 1-1, a).<sup>[36-37]</sup> Alternatively, a lactone derivative of the sugar (Scheme 1-1, b)<sup>[38-39]</sup> or the C1'-unprotected sugar (Scheme 1-1, c) is attacked by an organometallic nucleobase. In the first case, the resulting hydroxyl group is removed in a subsequent reduction (Scheme 1-1, b),<sup>[39]</sup> for the latter a Mitsunobu reaction is used for cyclization (Scheme 1-1, c). Cross couplings are a completely different strategy for *C*-nucleoside synthesis: e.g. a Heck reaction between a 1'-2' glycal and a Pd-activated aryl halide (Scheme 1-1, d).<sup>[40]</sup>



**Scheme 1-1.** Overview of common retrosynthetic strategies for *C*-glycosylation. Abbreviations: PG = protecting group, R = base derivative, M = metal, X = halogen.

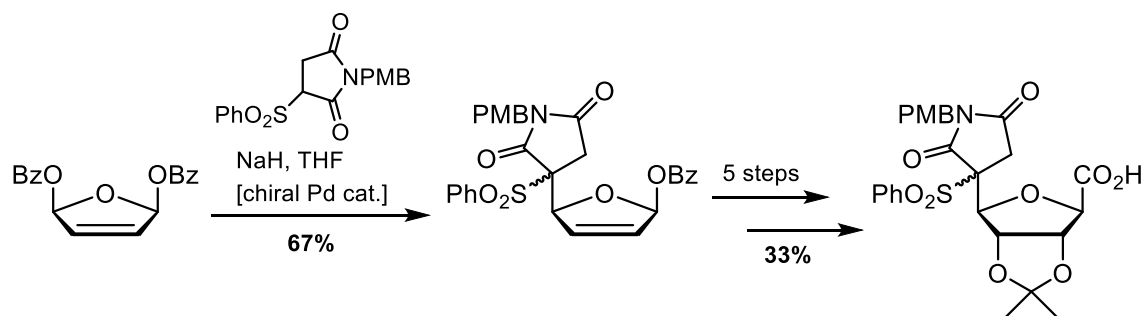
The key step for the synthesis of the salicylaldehyde nucleoside developed by *Clever et al.* is an example for the first approach.<sup>[12]</sup> A protected bromo base building block **2** was lithiated using *t*BuLi and subsequently transmetalated to a homocuprate (Scheme 1-2). This homocuprate is added to an  $\alpha$ -chloro deoxynucleoside derivative, Hoffer's chlorosugar **3**, and the desired  $\beta$ -nucleoside is obtained with 31% yield.



**Scheme 1-2.** C-glycosidation in the synthesis of the salicylaldehyde deoxynucleoside.

C-nucleosides containing (deoxy)ribose are usually prepared by the first three presented synthesis strategies (Scheme 1-1, a-c), despite the relatively low yields during C-glycosylation.<sup>[12, 41-42]</sup> Global deprotection of the compound after the glycosylation is often the only required transformation to obtain the nucleoside. Since the stereogenic centres at the sugar moiety (apart from C1') are preinstalled, additional reactions are mostly obsolete. Therefore, synthesis routes feature only few reaction steps.

Pd-catalyzed reactions like Tsuji-Trost<sup>[43]</sup> and Heck reactions<sup>[44]</sup> have been employed in the C-glycosylations as well. For example, the Tsuji-Trost reaction was applied in the total synthesis of L-showdomycin (Scheme 1-3).<sup>[43]</sup> After the glycosylation, five reaction steps were necessary to install the functional groups of the sugar. These approaches usually obtain higher yields for the C-glycosylation than alternative reactions without Pd, but are often less attractive as they increase the number of reaction steps.



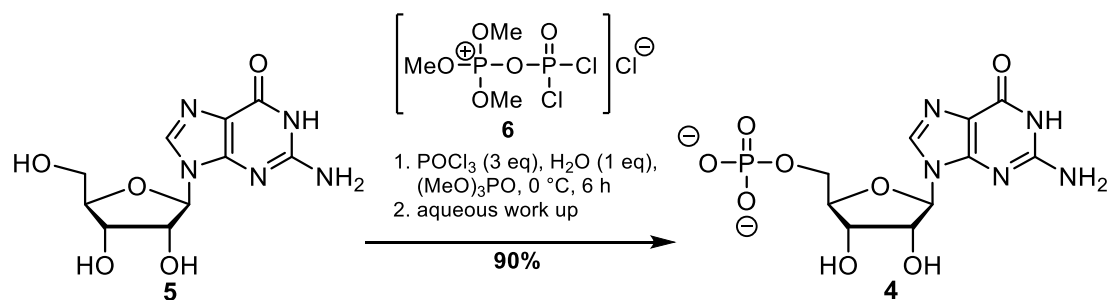
**Scheme 1-3.** Example of a Tsuji-Trost C-glycosylation using a chiral Pd catalyst.<sup>[43]</sup>

### 1.3 Chemical (Tri)phosphorylation Methods

Nucleotides are important compounds in the cellular metabolism, since they act as building blocks for enzymatic DNA and RNA synthesis. Moreover, they comprise chemically bound energy,<sup>[45]</sup> which is required for enzymatic reactions (ATP, UTP) and are involved in signal transduction<sup>[46]</sup> processes (ATP, GTP). Key to these functions is the charged phosphate group and the energy-rich phosphoanhydride bond. Whereas the most common eight nucleotides (dATP, dCTP, dGTP, dTTP, ATP, CTP, GTP, UTP) are

abundant in natural sources and can be obtained from isolated RNA and DNA of microorganism,<sup>[47]</sup> the preparation of modified nucleotides that at best have low natural abundance requires different strategies. Several synthetic methods exist, which provide access to nucleoside triphosphates with reasonable effort. However, low yields and tedious purifications are common problems. Despite major achievements, by *Ludwig*,<sup>[48]</sup> *Ruth*,<sup>[49]</sup> *Eckstein*,<sup>[33]</sup> *Hoard* and *Ott*<sup>[50]</sup> several decades ago, the chemical synthesis of triphosphate (derivates) still has potential for improvement. Numerous recent publications<sup>[51-54]</sup> demonstrate the ongoing need for improved methodology, as the number and diversity of applications involving (modified) nucleotides keeps growing. Some important chemical methods are highlighted in this chapter, since synthesis of artificial and modified triphosphates was essential for this work.

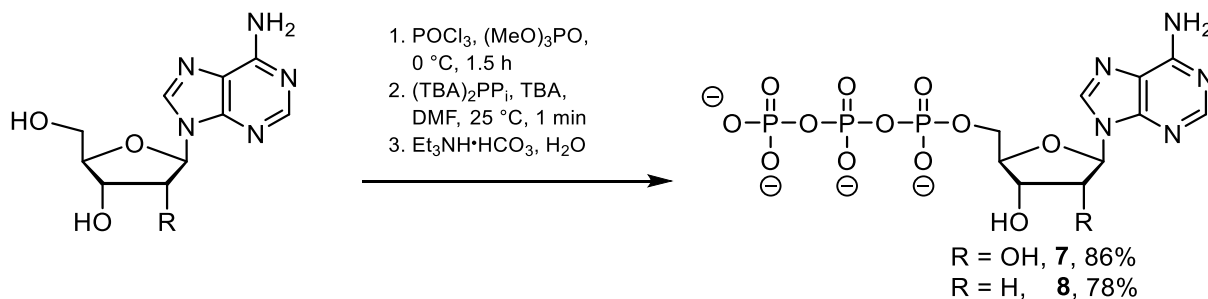
One of the first feasible chemical nucleotide synthesis was developed by *Yoshikawa* and coworkers in 1967. Guanosine monophosphate (**4**) was prepared from the corresponding nucleoside **5** using phosphorus oxytrichloride ( $\text{POCl}_3$ ) in trimethylphosphate (Scheme 1-4).<sup>[55]</sup> Applying contemporary analytical methods of that time, the protocol states 90% yield of the desired monophosphate and 5% additionally phosphorylated side products. In addition, 5'-monophosphates of inosine, adenosine, xanthosine, uridine and cytidine were obtained in yields of 80-91%. The excellent regioselectivity and yield is probably attributed to several factors. Addition of water generates  $\text{HCl}$  *in situ* and thereby decreases the nucleophilicity of the nucleoside after protonation and the effective amount of  $\text{POCl}_3$  present in the reaction. Trimethylphosphate as the solvent generates a homogenous solution and presumably reacts with  $\text{POCl}_3$  to form an active phosphorylation intermediate **6** similar to reactions of DMF with  $\text{POCl}_3$  or phosphoro-dichloridates.<sup>[56]</sup>



**Scheme 1-4.** Chemical monophosphate synthesis by *Yoshikawa et al.* and suggested phosphorylation intermediate **6**.<sup>[55]</sup>

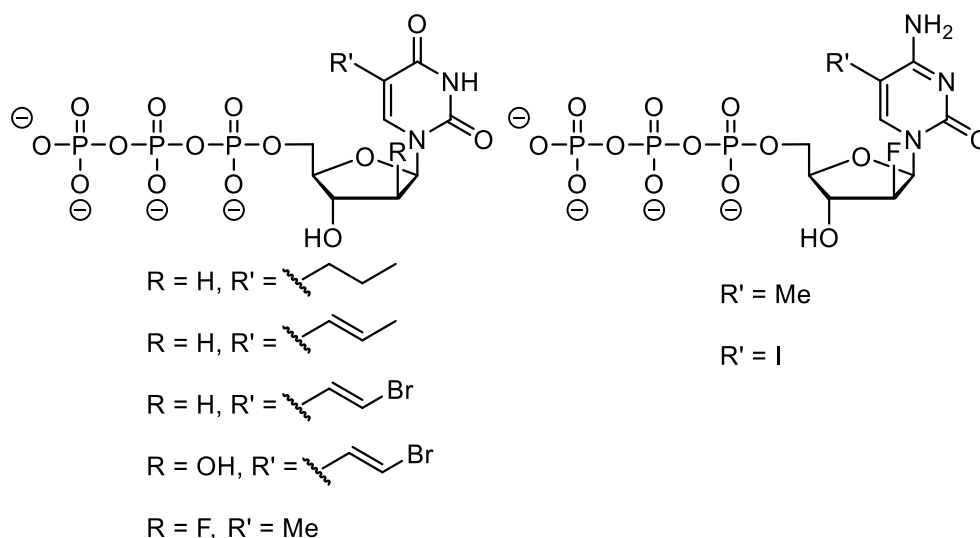
It was assumed that nucleoside dichlorophosphates are formed using this method, before hydrolysis during work-up. Apart from  $\text{H}_2\text{O}$ , this electrophilic intermediate could react with further nucleophiles. Consideration of this fact, prompted the concomitant development of nucleoside triphosphate syntheses by *Ludwig* and *Ruth*.<sup>[48-49]</sup> Based on the method from *Yoshikawa* and coworkers, nucleoside dichlorophosphates were prepared *in situ* and then reacted with bis(tri-*n*-butylammonium) pyrophosphate (( $\text{TBA}$ )<sub>2</sub> $\text{PP}_i$ ). The reaction was quenched by addition of triethylammonium bicarbonate before work-up. In this way, *Ludwig* obtained the 5'-triphosphates of adenosine **7** and 2'-deoxyadenosine **8** in 86% and 78% yield (determined by paper TLC), respectively (Scheme 1-5). In contrast to the *Yoshikawa* protocol, water

was omitted in the reaction to the nucleoside dichlorophosphate and addition of a base in the second step was found to improve the yield of the nucleoside triphosphates. Despite these changes, the selectivity for the 5'-triphosphate was excellent, as less than 0.5% of the 3'-analog from dATP were stated.<sup>[48]</sup>



**Scheme 1-5.** Chemical triphosphate synthesis by Ludwig.<sup>[48]</sup> TBA = tributylamine,  $(\text{TBA})_2\text{PP}_i$  = bis(tri-*n*-butylammonium) pyrophosphate.

At the same time *Ruth* and *Cheng* developed a protocol which was practically identical to the procedure published by *Ludwig*.<sup>[49]</sup> Several thymidine and 2'-deoxycytidine analogs were converted into the corresponding 5'-triphosphates (Figure 1-3) and studied for their ability to inhibit human polymerases ( $\alpha$  and  $\beta$ ) and *Herpes simplex* virus polymerases. Isolated triphosphate yields of 8-46% (mainly 20-35%) were obtained for seven different nucleoside analogs in 95-99% purity (according to HPLC). Lack of an additional base during the second reaction step was the only difference to the method from Ludwig. The small set of artificial nucleotide analogs already indicated a wide applicability of the chemical triphosphorylation method.

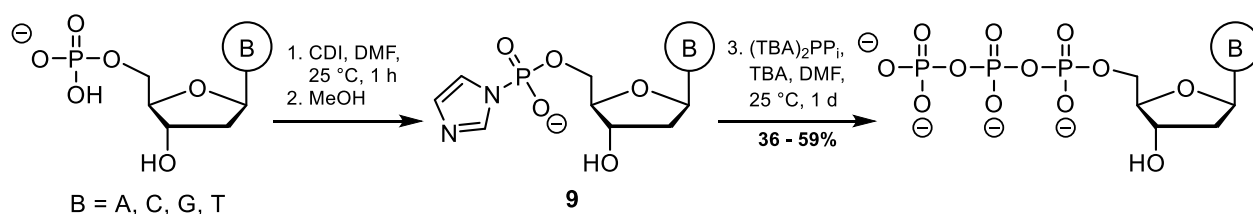


**Figure 1-3.** Nucleotide analogs synthesized by *Ruth* and *Cheng* from the corresponding nucleosides. Reaction conditions: i)  $\text{POCl}_3$ ,  $(\text{MeO})_3\text{PO}$ ,  $-10\text{ }^\circ\text{C}$  to rt, 3–15 h, ii)  $(\text{TBA})_2\text{PP}_i$ , DMF,  $-10\text{ }^\circ\text{C}$  to rt, 2–3 h, iii)  $\text{Et}_3\text{N}_{(\text{aq})}$  to pH 7.

Over the intervening years, numerous other artificial nucleotides were prepared by this “one-pot, three step” method.<sup>[8, 13, 42, 57-58]</sup> Despite its popularity, this nucleoside triphosphate synthesis method is not

universally applicable. For some substrates no 5'-triphosphate or only low yields were obtained, probably due to regio- and chemoselectivity issues.<sup>[13]</sup>

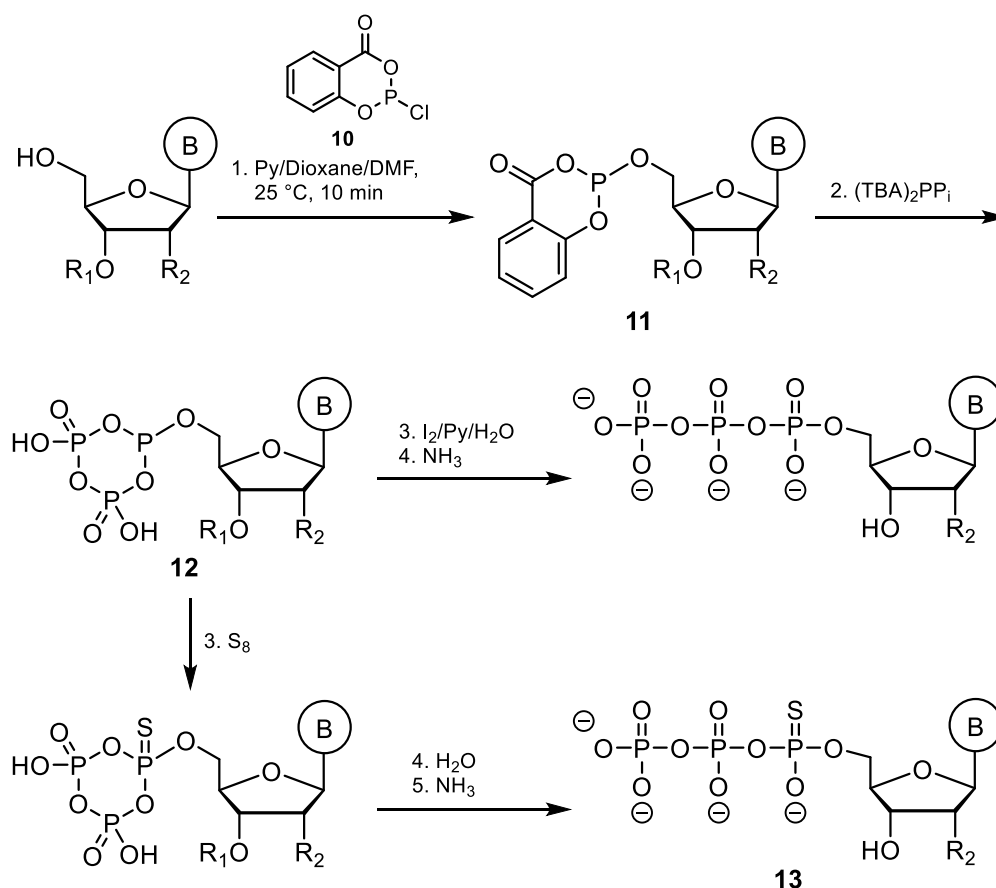
Preparation of the 5'-monophosphate (including isolation) and subsequent reaction to the triphosphate circumvents some of the problems encountered in the one-pot method. In 1959, reaction of 1,1'-carbonyldiimidazole (CDI) with nucleotide monophosphates to the corresponding imidazolides in mostly quantitative yield was published. It was noted that the resulting phosphorimidazolides represent valuable intermediates, which readily react with phosphates, alcohols, amines and carboxylic acids.<sup>[59]</sup> This enabled the synthesis of 5'-diphosphates and even dinucleotide phosphates from monophosphates.<sup>[60]</sup> By addition of bis(tri-*n*-butylammonium) pyrophosphate to the *in situ* formed phosphorimidazolide **9**, triphosphates were prepared from the four 2'-deoxynucleoside-5'-monophosphates in 36-59% isolated yield (Scheme 1-6) by *Hoard and Ott*.<sup>[50]</sup> Even the 5'-phosphate of di- and trinucleotides was converted to the corresponding 5'-triphosphate in 20-70% isolated yield.<sup>[50]</sup>



**Scheme 1-6.** Activation of nucleotide monophosphates to phosphorimidazolides **9** and reaction to triphosphates.

The formation of 2',3'-cyclic carbonates with ribonucleoside-5'-monophosphates is one disadvantage of CDI which reduces the triphosphate yield.<sup>[61]</sup> A sulfonyl imidazolium derivative, as an alternative to CDI, provides isolated triphosphate yields of 84-90% starting from ribonucleoside monophosphates.<sup>[53]</sup> The suggested reaction mechanism of this more recent approach is discussed in a subsequent chapter (Scheme 3-22, p. 92).

A completely different strategy, relying on phosphites as electrophilic intermediates, was developed by *Ludwig and Eckstein* in 1989.<sup>[33]</sup> Using the salicylphosphite reagent **10**,<sup>[62]</sup> a nucleoside phosphite **11** is formed within 10 min, which results in a cyclic phosphate intermediate **12** upon addition of bis(tri-*n*-butylammonium). Oxidation and hydrolysis of this intermediate gives the corresponding triphosphate (Scheme 1-7). The proposed reaction mechanism was supported by <sup>31</sup>P NMR of the crude reaction intermediates.<sup>[33]</sup> To achieve good yields and sufficient 5'-selectivity, protection of the 3'-hydroxyl group is required. *In situ* deprotection by aqueous ammonia after triphosphate formation yielded, e.g. thymidine 5'-triphosphate from 3'-acetylthymidine in 72% yield. Notably, all reaction steps are performed in one flask, without isolation of intermediates and provide moderate to good yields even of modified nucleotides (22-72%). A slight modification of the oxidation protocol, using S<sub>8</sub> instead of I<sub>2</sub>, gives 5'-O-(1-thiotriphosphate)s **13** as *R/S*-diastereoisomer mixtures. Incorporation of these  $\alpha$ -sulfur-modified triphosphates **13** by polymerases produces DNA and RNA molecules containing phosphorothioate backbones. This improves the resistance of these macromolecules towards nucleases.<sup>[63]</sup>

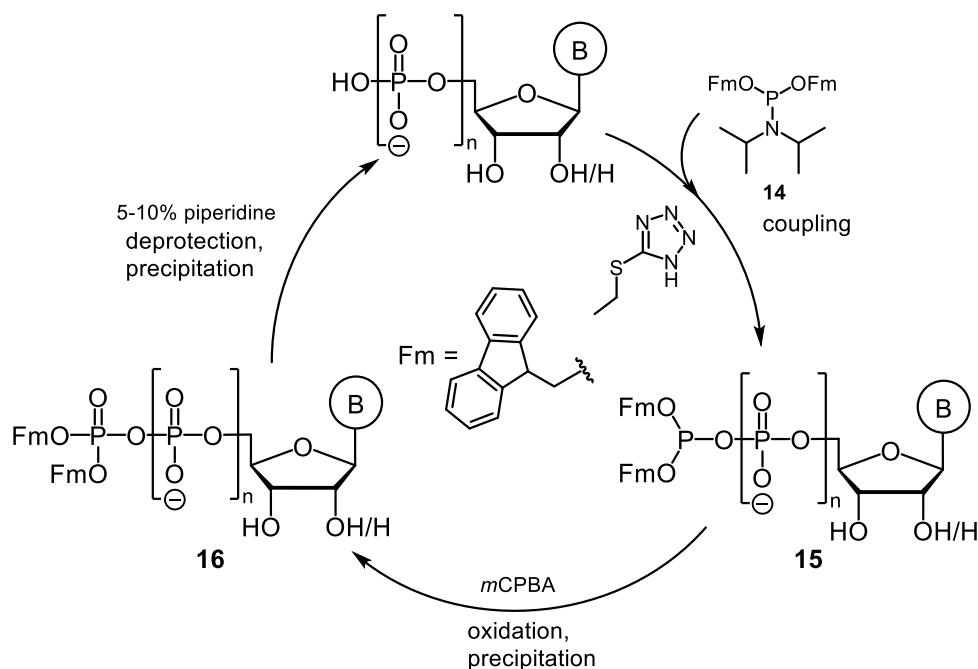


**Scheme 1-7.** Chemical triphosphorylation method by *Ludwig* and *Eckstein*.<sup>[33]</sup> R<sub>1</sub> = ester protecting group, R<sub>2</sub> = H, or OR (OH in the triphosphate product). B = nucleobases A, C, G, T, U.

A key variation of the protocol from *Ludwig* and *Eckstein*, is based on a changed order of reagent addition. Thereby, protecting group-free synthesis of all eight major nucleotides and also an artificial nucleotide was achieved from the corresponding nucleosides in yields ranging from 10-50%.<sup>[51, 64]</sup> The selectivity ratio for triphosphorylation of the 5'-hydroxyl over the 3'- or 2'-positions was determined to 85:15.<sup>[64]</sup> Detailed discussion and application of this alternative protocol can be found in chapter 3.1, Scheme 3-4.

The phosphoramidite group is the synthetic equivalent of an electrophilic phosphate, therefore its application in chemical phosphorylations seems obvious. In analogy to the recurring synthesis cycle in oligonucleotide synthesis, an iterative phosphorylation protocol was developed by *Jessen* and coworkers.<sup>[54, 65]</sup> First, a fluorenylmethyl-protected P(III) amidite **14** is coupled to a nucleoside 5'-monophosphate in 10 min (Scheme 1-8). Notably, the reaction is chemoselective under the applied conditions and no protecting groups are necessary on the nucleobase and the sugar. Moreover, anhydrous solvents and reagents are obsolete, thus simplifying the protocol. The resulting mixed P(III)-P(V) anhydride **15** is oxidized to the P(V)-form **16** by *m*CPBA in 1 min and the product is precipitated from solution. Treatment with 5-10% piperidine deprotects the terminal phosphate group and yields the 5'-diphosphate after an additional precipitation. The reaction cycle can be started again by coupling of a phosphoramidite, oxidation and deprotection to the triphosphate.<sup>[54]</sup> This iterative procedure provided the

diphosphates of the canonical nucleotides in 75-93% isolated yield from the corresponding monophosphates. The whole procedure is complete within 30 min and was also used to prepare the tetraphosphate from ATP in 50% isolated yield, thus demonstrating its general applicability. In addition, this method was successfully transferred to a solid-phase protocol, preparing dGTP from commercially available protected 2'-deoxyguanosine CPG support.<sup>[54]</sup>



**Scheme 1-8.** Principle of the iterative oligophosphate synthesis developed by *Jessen* and coworkers.<sup>[54]</sup>

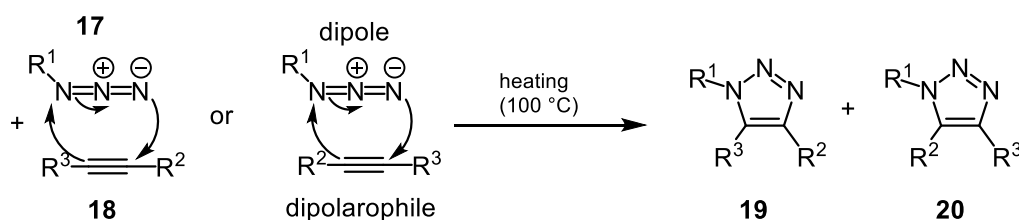
## 1.4 Cu(I)-Catalyzed Alkyne-Azide Cycloaddition

The concept of “click” chemistry was introduced in 2001 by *Kolb, Finn* and *Sharpless*. So-called “click” reactions are used to connect two molecular building blocks selectively and are characterized by high yields under ambient conditions (water-tolerant) in the presence of many functional groups.<sup>[66]</sup> The bioorthogonal Cu(I)-catalyzed alkyne/azide cycloaddition (CuAAC) is arguably the most widely used click reaction.<sup>[67]</sup>

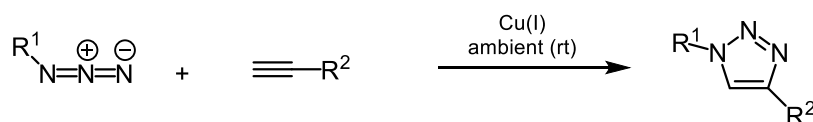
Thermally and light-induced reactions between alkynes and azides have been known since 1893, when the first synthesis of a 1,2,3-triazoles was described by *Arthur Michael*.<sup>[68]</sup> Due to detailed systematic studies of 1,3-dipolar cycloadditions in the mid-20<sup>th</sup> century by *R. Huisgen*, the alkyne-azide cycloaddition became known as the Huisgen reaction.<sup>[69-70]</sup> Based on his investigations, *Huisgen* proposed a concerted pericyclic mechanism for the reaction between a 1,3-dipole (here: azide **17**) and a dipolarophile (here: alkyne **18**). A stepwise mechanisms for the thermally induced reaction (Scheme 1-9, top) was excluded. Despite its applicability for a variety of substrates in the synthesis of five-membered heterocycles, the need for elevated temperature (about 100 °C) and the resulting mixture of 1,4 **19** and 1,5 regioisomers **20**, have limited the practical scope of the reaction. In 2002, two research groups independently discovered that Cu(I) catalysis provides selectively 1,4-disubstituted 1,2,3-triazoles from alkynes and azides using ambient

conditions (Scheme 1-9, bottom).<sup>[71-72]</sup> Various alkynes and azides were efficiently reacted to triazoles in good to excellent yields in water containing solvent mixtures and in the presence of many functional groups.<sup>[71-72]</sup> The simplicity, efficiency and the broad applicability of the CuAAC has led to more than 1000 research articles within the first decade after the original discovery by the groups of *Fokin*, *Sharpless* and *Meldal*.<sup>[67]</sup>

Huisgen's 1,3-dipolar cycloaddition



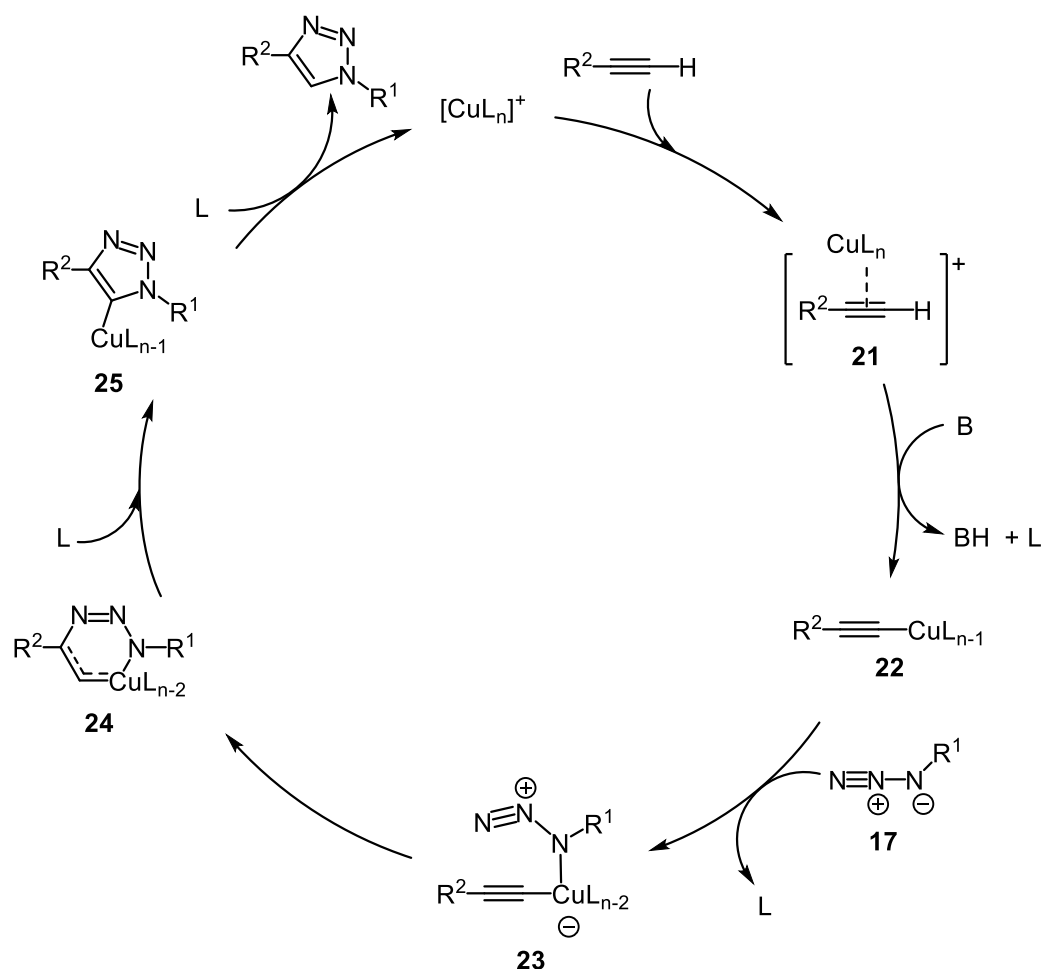
CuAAC reaction



**Scheme 1-9.** Overview of the Huisgen reaction and the Cu(I)-catalyzed version (CuAAC).

In contrast to the concerted Huisgen reaction, the Cu(I)-catalyzed version is a stepwise non-concerted reaction.<sup>[71]</sup> Kinetic studies and DFT calculations were performed to further investigate the mechanism of the CuAAC.<sup>[73-76]</sup> The results suggest a complex multi-step mechanism that starts by initial formation of a Cu(I)-alkyne  $\pi$ -complex **21** (Scheme 1-10). This complex lowers the  $pK_a$  of the terminal alkyne by 10 units, which allows more facile deprotonation and results in a Cu(I)-acetylide **22**. Then, the azide **17** binds to the copper to form intermediate **23** by replacing one of the ligands (here:  $H_2O$  or MeCN assumed). The distal nitrogen of the azide is attacked, generating a six-membered metallacycle **24** containing a Cu(III) species. DFT calculations suggest that this intermediate is very important for the rate acceleration of the catalyzed reaction in comparison to the uncatalyzed concerted reaction.<sup>[75]</sup> Subsequent ring contraction to the triazolyl-copper **25** is accompanied by reduction of Cu(III) to Cu(I) and ligand addition. The following proteolysis releases the triazole and completes the catalytic cycle.





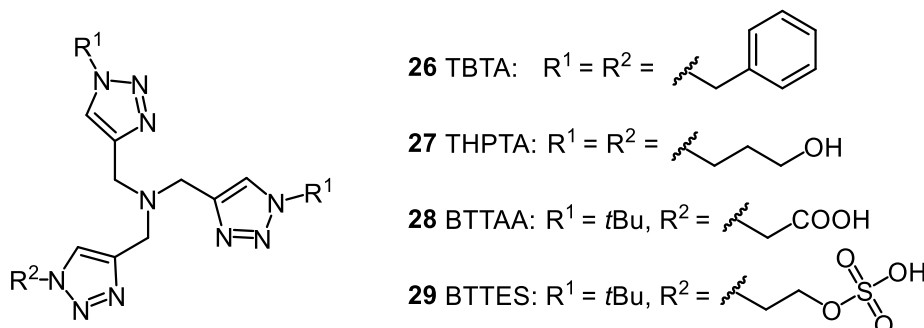
**Scheme 1-10.** Proposed mechanism of the CuAAC adapted according to Himo *et al.* and Rostovtsev *et al.*<sup>[71, 75]</sup> B = generic base, L = generic ligand,  $\text{R}^1$  and  $\text{R}^2$  = any residue except H.

Experimental data provided evidence for the transition of the alkyne-azide copper complex **23** to the triazolyl-copper **25** as the rate-determining step of the CuAAC reaction.<sup>[77]</sup> Alternative proposals for the catalytic cycle involve bimetallic structures and more complicated mechanisms.<sup>[74, 76, 78]</sup>

One key finding, which has contributed to the wide applicability of the CuAAC, was a reaction rate increase in the presence of some amine-containing compounds.<sup>[79]</sup> These molecules possess polydentate nitrogen donors which can become ligands of Cu(I), thus preventing the formation of unreactive Cu(I) aggregates<sup>[80]</sup> and disproportionation to Cu(0) and Cu(II).<sup>[81]</sup> Examples of favorable ligands are TBTA **26**,<sup>[79-80]</sup> THPTA **27**<sup>[82]</sup> and BTAA **28**,<sup>[83]</sup> which can increase the reaction rate up to 10-100-fold as compared to the ligand-free CuAAC. These ligands are azide-alkyne click products and thus contain triazoles. Moreover, they possess a tris(triazolylmethyl)amine-motif and only differ in the residues of the former azides (Figure 1-4).

Especially bioconjugations profit from ligand-accelerated CuAACs, as high amounts of copper salts have a toxic effect and can hinder *in vivo* applications.<sup>[84]</sup> For example the ligands BTAA **28** and BTES **29** were successfully applied for CuAAC-promoted labeling in live zebrafish embryos.<sup>[83-84]</sup> Due to the enormous rate-accelerating effect of the ligand molecules, only 40  $\mu\text{M}$   $\text{CuSO}_4$  precatalyst was needed

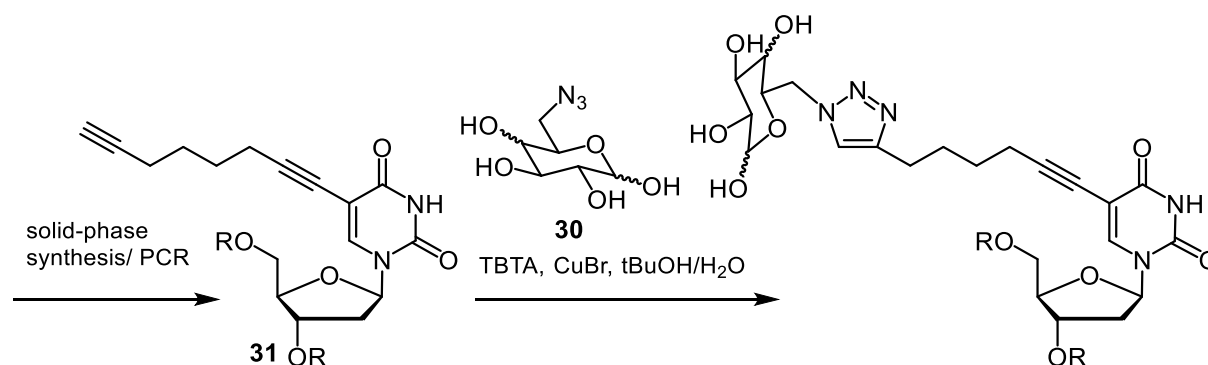
(instead of 500  $\mu\text{M}$ ), which did not cause developmental defects in the click-labeled zebrafish embryos.<sup>[83]</sup> Also compounds without nitrogen donors can have a rate-accelerating effect on the CuAAC, like e.g. carboxylates and triphenyl phosphine  $\text{PPh}_3$ .<sup>[85]</sup> Furthermore, solvent ligand effects have been reported.<sup>[86]</sup>



**Figure 1-4.** Structural formulas of common ligands used in CuAAC. TBTA = tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine (**26**), THPTA = tris[(1-hydroxypropyl-1H-1,2,3-triazol-4-yl)methyl]amine (**27**), BTAA = 2-[4-[(bis[(1-tert-butyl-1H-1,2,3-triazol-4-yl)methyl]amino)methyl]-1H-1,2,3-triazol-1-yl]acetic acid (**28**), BTES = 2-[4-[(bis[(1-tert-butyl-1H-1,2,3-triazol-4-yl)methyl]amino)methyl]-1H-1,2,3-triazol-1-yl]ethyl hydrogen sulfate (**29**).

Using CuAAC for the modification of oligonucleotides has been particularly fruitful.<sup>[87-89]</sup> Cu(I)-stabilizing ligands were important to the application of CuAAC with DNA, as non-chelated copper ions can damage DNA by strand scission.<sup>[90]</sup> Nucleotides and DNA were prepared containing alkyne groups, which were subsequently functionalized with azide-containing reporter molecules by CuAAC.<sup>[87]</sup> This approach enabled the efficient and selective introduction of many modifications, like e.g. fluorophores,<sup>[91]</sup> sugars,<sup>[92-93]</sup> dendrimers,<sup>[94]</sup> affinity-tags,<sup>[95]</sup> and aldehydes for metallization.<sup>[96]</sup>

Pioneering work from the *Carell* group established the CuAAC as efficient method to post-synthetically modify DNA.<sup>[91,96]</sup> Therefore, alkyne labels were introduced in the 5-position of the pyrimidine nucleobase. The phosphoramidites of the alkyne-uridine analog were synthesized and applied in the solid-phase synthesis of alkyne-labeled DNA strands. Varying label densities and two alkyne linker length were studied for their ability to undergo efficient CuAAC with fluorophore- and sugar azides **30** (Scheme 1-11). Up to six consecutive labels were introduced by click chemistry for the uridine analog **31** containing an alkyne on a flexible spacer.<sup>[91,96]</sup> Notably, the corresponding alkyne-modified deoxynucleoside triphosphates (uridine and cytosine analogs) were even accepted as a substrate in polymerase chain reactions (PCR). Sequencing proved that the uridine and cytosine derivatives were faithfully incorporated opposite a templating A and G, respectively. More than 800 alkyne modifications were introduced into a 2000 bp long PCR product, of which at least 95% reacted with a sugar azide in a subsequent click reaction.<sup>[93]</sup>



**Scheme 1-11.** CuAAC between an alkyne-modified DNA strand and a glucosyl/galactosyl azide **30** as performed by Gierlich *et al.* [91, 93] R = oligonucleotide

Several methods have been developed to provide DNA with a “clickable” backbone.<sup>[97-100]</sup> For example an alkynyl phosphinoamidite reagent was developed, which allowed introduction of an alkynyl-phosphonate by solid-phase synthesis and subsequent modification by CuAAC.<sup>[98]</sup> 3'-Azide and 5'-alkyne-modified nucleosides were introduced into oligonucleotides by solid-phase synthesis and post-synthetic activated ester modification.<sup>[97, 100]</sup> The efficiency of CuAAC was then exploited to selectively join azide- with alkyne-containing DNA strands. This form of “chemical ligation” was even capable of joining non-templated DNA strands with good efficiency. If both modifications were within a single strand, even non-templated circularization was achieved at low DNA concentrations (0.4  $\mu\text{M}$ ).<sup>[97]</sup> Remarkably, the templates generated by chemical ligation, which contain a triazole in the backbone, can be used for PCR<sup>[100]</sup> and even for *in vivo* transcription.<sup>[101]</sup>

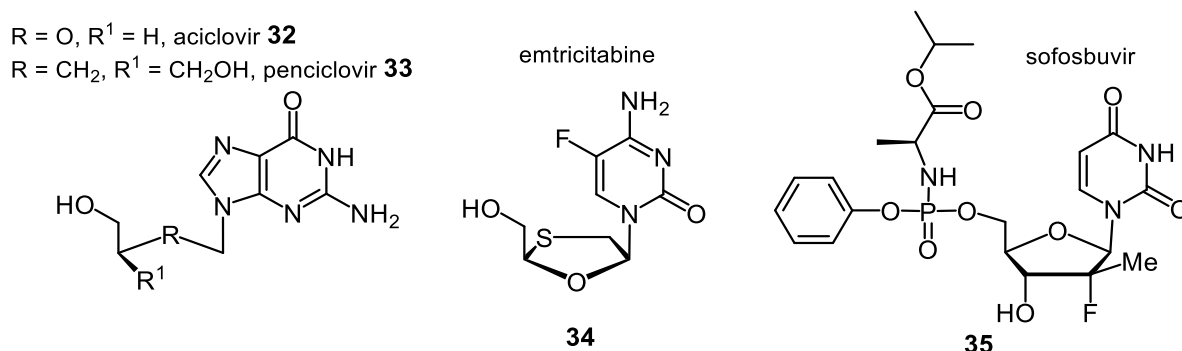
As many molecular-based diagnostic methods require photometrical readouts, fluorophores are the most common modification in reporter molecules.<sup>[102]</sup> The most important applications of fluorescently-labeled (oligo)nucleotides are arguably sequencing approaches<sup>[103]</sup> and probes for fluorescent *in situ* hybridization (FISH) in genome analysis.<sup>[104]</sup> Nucleotides with fluorophores on the terminal phosphates are suitable substrates in single-molecule real-time sequencing methods.<sup>[3,7]</sup> Despite numerous applications of the CuAAC on oligonucleotides, it had not been used for the synthesis of phosphate-modified nucleotides.

## 1.5 Application of Artificial Nucleosides and Nucleotides

### 1.5.1 Therapeutic Nucleosides and Nucleotides

The first nucleosides and nucleotides that were prepared by chemical synthesis were essential for the structural elucidation of natural nucleosides.<sup>[20]</sup> In order to study their biological role, a sufficient supply of these compounds was necessary. By developing the chemical methods of nucleoside chemistry, several non-natural nucleosides were prepared as model substrates.<sup>[105]</sup> Thereafter, the first non-natural nucleosides and nucleotides were studied for their ability to be accepted as substrates for nucleoside-converting enzymes *in vitro*.<sup>[106]</sup> This provided insight into biocatalysis of these enzymes and ultimately yielded candidates for antiviral therapy. Several artificial nucleosides have become efficacious antiviral agents for

treatment of retroviral diseases. The most common agents arguably are acyclovir **32**<sup>[107]</sup> and penciclovir **33**<sup>[108]</sup> used in labial *Herpes simplex* infections (Figure 1-5). Notably, acyclovir is non-toxic to normal cells and only displays its antiviral effect in infected cells.<sup>[109]</sup> Even severe retroviral diseases like AIDS and hepatitis infections are treated by nucleoside analogs. One example is emtricitabine **34** (Figure 1-5), which is a reverse transcriptase inhibitor used for the treatment against human immuno deficiency virus.<sup>[1]</sup> Many of these compounds are antimetabolites of the natural nucleosides (and nucleotides) and terminate the viral RNA and DNA polymerisation upon incorporation of the nucleotide analog.



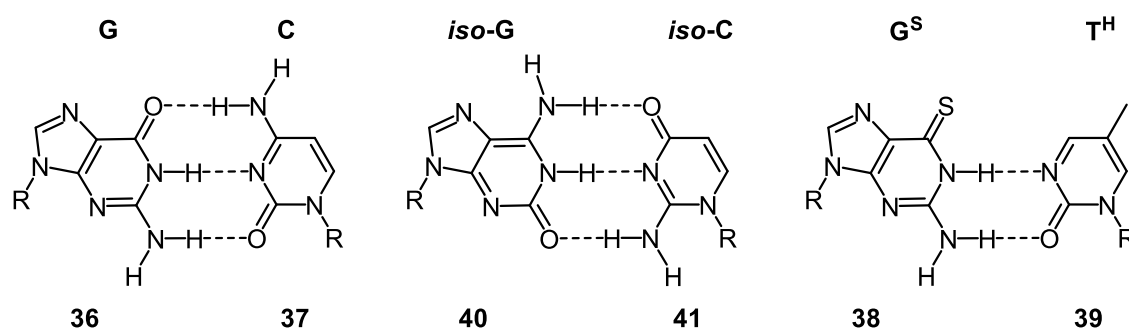
**Figure 1-5.** Structural formulas of selected nucleoside analogs used in antiviral therapy.

It is noteworthy that most of these nucleoside drugs must be transformed into the corresponding triphosphate within the cell in order to display an inhibitory effect. Especially the first phosphorylation step to the monophosphate can be slow and thus lower the efficacy of the drug, as the intracellular triphosphate amount becomes limiting.<sup>[110]</sup> In order to circumvent this problem, the more recently developed sofosbuvir **35** (Figure 1-5) represents a monophosphate prodrug with good bioavailability, which is quickly reacted to the corresponding triphosphate intracellularly.<sup>[111]</sup> This uridine analog **35** is used in the treatment against hepatitis C viral infections. Moreover, artificial nucleosides are used for the treatment of Ebola virus disease (BCX4430)<sup>[112]</sup> and cancer (e.g. gemcitabine),<sup>[2]</sup> thus demonstrating the importance of this compound class for human health.

### 1.5.2 Unnatural Base Pairs – Alternative Hydrogen Bonding

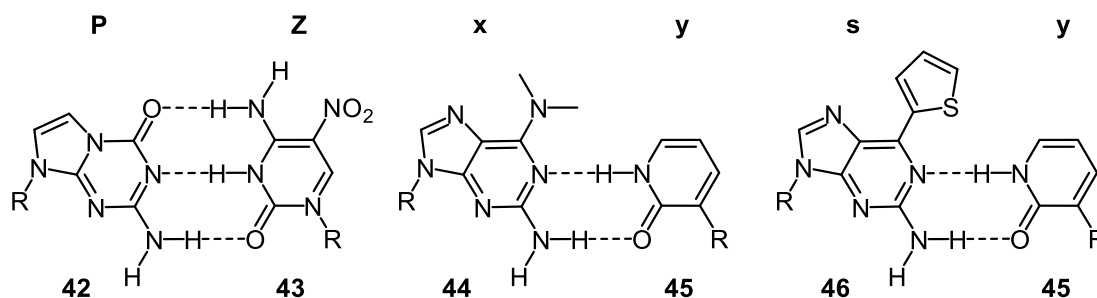
Instead of inhibiting the nucleotide-consuming enzymes, the creation of artificial nucleotides, which cooperate with the enzyme machinery like a natural nucleotide represents an even greater challenge. Since natural DNA consists of four major bases which form two different base pairs, a triplet codon is necessary to provide  $4^3 = 64$  combinations. Living organisms encode with 61 codons 20 canonical amino acids and they utilize three stop codons within these 64 combinations.<sup>[113]</sup> By adding a single base pair to the natural bases the number of potential codons could be increased to  $6^3 = 216$ , thus providing new codons for non-canonical amino acids. This expansion of the genetic code is one major goal in synthetic biology and has the potential to provide access to novel biocatalysts by semi-synthetic organisms in the future.<sup>[4]</sup>

More than 50 years ago, the idea of an unnatural base pair was proposed, which could provide DNA with novel functionality and expand the genetic code.<sup>[114]</sup> In the late 1980s, the first studies on unnatural base pairs were performed in the groups of *Rappaport*<sup>[115]</sup> and *Benner*,<sup>[116]</sup> using the concept of alternative hydrogen bonding patterns. These unnatural nucleobases represented only slight modifications of the natural G:C pair **36:37** (Figure 1-6). *Rappaport* observed that his 6-thioguanine:5-methyl-2-pyrimidinone pair ( $G^S:T^H$ , **38:39**) was about as stable as a natural A:T base pair.<sup>[115]</sup> Subsequent studies involving the large Klenow fragment (KF) polymerase I from *Escherichia coli* (*E. coli*) observed mutation of  $G^S:T^H$  to G:C.<sup>[117]</sup> *Benner* and coworkers demonstrated sufficient fidelity of *in vitro* replication and transcription for their *iso*-G:*iso*-C base pair **40:41**.<sup>[118]</sup> Moreover, ribosomal translation of this unnatural base pair into a non-canonical amino acid was achieved *in vitro*.<sup>[119]</sup> These early efforts represented a first proof-of-concept that an expansion of the genetic code by unnatural base pairs is possible.



**Figure 1-6.** Structural formula of the natural G:C base pair compared to the unnatural *iso*-G:*iso*-C and  $G^S:T^H$  base pairs, which interact *via* an alternative hydrogen bonding pattern. R = Sugar and phosphate residues.

Despite these outstanding results, the *iso*-G:*iso*-C base pair **40:41** has several shortcomings. The tautomeric enol-form of the *iso*-G **40** (*iso*-G\*) forms a mispair with T, the deamination product of *iso*-C **41** is misinserted opposite of A and finally the transcription is inefficient.<sup>[120]</sup> By replacing the natural thymidine triphosphate (dTTP) with a 2-thiothymidine variant, the *iso*-G\*:T misinsertion is destabilized, while the A:T pair remains unaltered, thus improving the fidelity of the *iso*-G:*iso*-C replication. This improvement is sufficient to allow PCR amplification of DNA containing the unnatural base pair.<sup>[121]</sup> A fidelity of 98% was determined for each replication of the *iso*-G:*iso*-C pair **40:41** under these conditions, leading to a loss of 32.2% ( $1-0.98^{20}$ ) of the unnatural base pairs after 20 PCR cycles. Redesign of the *iso*-G:*iso*-C **40:41** recently resulted in the P:Z pair **42:43**, which does not tautomerize and which is formed with better selectivity (Figure 1-7). The fidelity in PCR amplification was improved to 99.8% per cycle<sup>[122]</sup> and even transcription and reverse transcription was performed.<sup>[123]</sup>

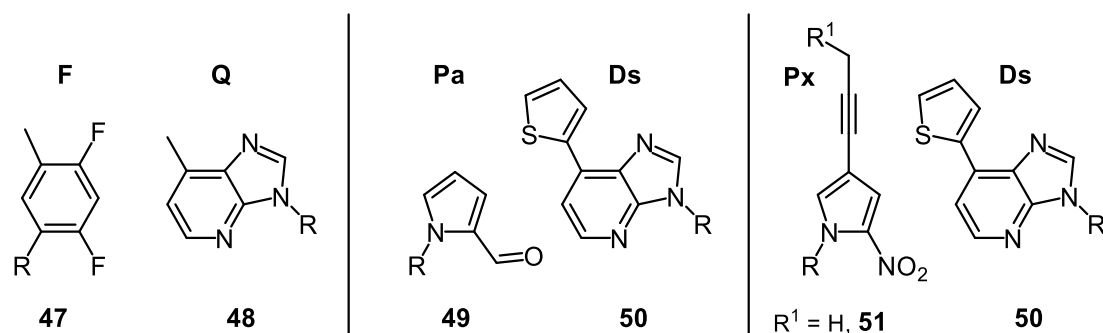


**Figure 1-7.** Structural formula of the P:Z, x:y and s:y pair. R = Sugar and phosphate residues.

Hirao and coworkers developed artificial base pairs, which also relied on hydrogen bonding for interaction. The design especially of the initial x:y pair **44:45** (Figure 1-7) closely resembled the  $\text{G}^{\text{S}}:\text{T}^{\text{H}}$  **38:39** (Figure 1-6) reported by *Rappaport* a decade earlier. Therefore, it was not so surprising that the x:y base pair **44:45** was mutated during replication by polymerases.<sup>[58, 124]</sup> In order to decrease the mispair formation, a sterically more demanding thiophenyl group was installed at the 6-position of the purine nucleobase **46**.<sup>[125]</sup> This modification improved the selective enzymatic pairing and transcription of the s:y pair **46:45** (Figure 1-7). Although transcription of s **46** into y **45** was faithfully possible, transcription of y **45** into s **46** was not satisfying. Hence, an artificial tRNA had to be prepared by a combination of chemical synthesis and enzymatic ligation, containing s **46** in the anticodon loop, to circumvent the selectivity issue. This allowed *in vitro* transcription and translation of the s:y unnatural base pair **46:45** into a 3-chlorotyrosine containing protein.<sup>[126]</sup>

### 1.5.3 Hydrophobic Unnatural Base Pairs

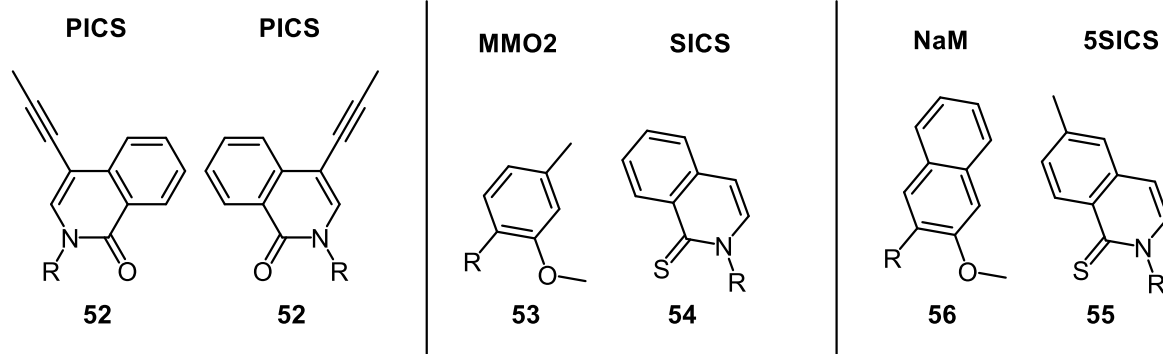
Already in 1958, experimental and theoretical results by *Sturtevant et al.* indicated that hydrogen bonds are not the sole force of DNA stability.<sup>[127]</sup> Several factors have been identified which contribute to the non-hydrogen bonding interactions. Among these, geometrical selection, the hydrophobic effect and packing forces between aromatic systems, also known as  $\pi$ -stacking, have been identified to be most important.<sup>[4]</sup> In 1997, *Kool* and coworkers prepared the thymidine analog, 2,4-difluorotoluene nucleoside **F 47**, which was incorporated selectively opposite a templating adenine by KF polymerase *in vitro*.<sup>[128]</sup> In addition, an isosteric adenine variant **Q 48**, lacking all interstrand hydrogen bond donating and accepting groups was developed. It was demonstrated that **Q 48** was an efficient template for **F 47** insertion.<sup>[11]</sup> The absence of any hydrogen bond between the F:Q pair **47:48** (Figure 1-8) clearly demonstrated that alternative interactions are sufficient for selective pairing and that hydrogen bonding is not essential to form base pairs.<sup>[11, 129]</sup>



**Figure 1-8.** Structural formulas of unnatural base pairs from the *Kool* group (F:Q) and from the *Hirao* group (Pa:Ds and Px:Ds). R = Sugar and phosphate residues. R<sup>1</sup> = HNCO(CH<sub>2</sub>)<sub>5</sub>-NH<sub>2</sub>, **51\*** was used to introduce modifications after PCR.

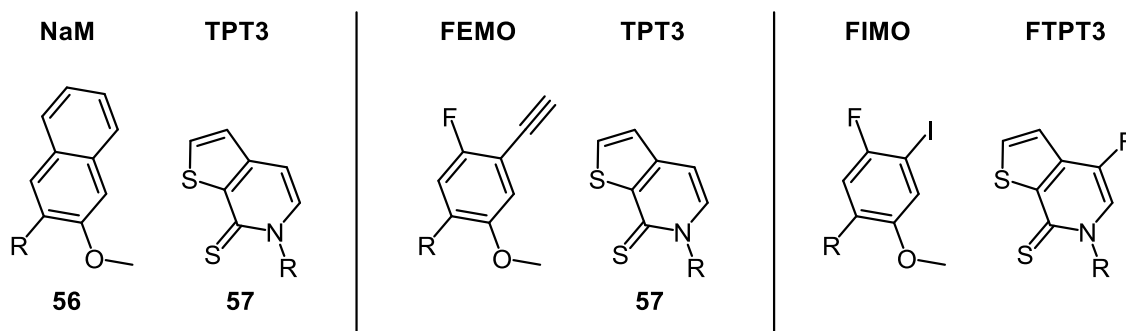
Using the concept of predominately hydrophobic pairing, *Hirao* and coworkers developed the Pa:Ds unnatural base pair **49:50** (Figure 1-8). PCR amplification with a fidelity of more than 99% per cycle was achieved, when  $\gamma$ -amidotriphosphates of Ds and dA were used and transcription succeeded with less than 5% mutation of the Pa:Ds **49:50** pair.<sup>[130]</sup> Replacing the aldehyde of the Pa **49** with a nitro group and introducing a conjugated alkyne resulted in the Px:Ds pair **51:50** (Figure 1-8). This variant improved the fidelity during PCR to more than 99.9% per cycle and allowed introduction of a fluorophore after enzymatic amplification if a modified Px **51\*** was used.<sup>[10]</sup>

The *Romesberg* group has extensively studied the concept of unnatural base pairs, which rely on hydrophobic effect and  $\pi$ -stacking.<sup>[4]</sup> Early studies majorly focussed on bicyclic self-pairs, featuring naphthyl and indole derivatives. Several potential unnatural base pairs were identified, which were incorporated by polymerases at rates one to three orders of magnitude slower than the natural nucleotides and possessed acceptable selectivities.<sup>[131]</sup> However, synthesis after the incorporation of the unnatural bases was inefficient. NMR structure of a DNA containing the PICS-PICS self-pair **52:52** (Figure 1-9) revealed that the large aromatic areas pair in an intercalative manner, which might misplace the 3'-OH terminus for the subsequent nucleotide addition.<sup>[132]</sup> In order to circumvent this problem, monocyclic aromatic systems, like pyridine and benzene derivatives were studied as unnatural base pair candidates.<sup>[44, 133-135]</sup> Some of these candidates were incorporated and extended by DNA polymerase with reasonable efficiency.<sup>[134]</sup> The key observations from these studies were conflicting with a rational design approach: Hydrophobic interactions favor fast formation of the unnatural base pair, but hinder further extension. Minor groove hydrogen bond acceptors favor subsequent elongation, yet decrease the efficiency of initial unnatural base pair formation. Therefore, a screening effort comprising more than 3600 possible artificial base pairs was undertaken to investigate formation and extension of unnatural base pairs. This screen found the MMO2:SICS pair **53:54** (Figure 1-9) to perform best.<sup>[136]</sup> Moreover, the two nucleobases possess minor groove hydrogen bond acceptors *ortho* to the glycosidic bond. Both hydrogen bond acceptors (methoxy and thiocarbonyl) are rather hydrophobic, thus combining the “contradictory” properties required for efficient enzymatic replication.



**Figure 1-9.** Structural formulas of selected unnatural base pairs from the Romesberg group. R = Sugar and phosphate residues.

Introduction of a methyl group to the SICS, further improved the unnatural base pair and resulted in the 5SICS base **55**, which still paired with MMO2 **53**. This MMO2:5SICS pair **53:55** was replicated by several different DNA polymerases.<sup>[137]</sup> The NaM:5SICS pair **56:55** (Figure 1-9) was obtained from further optimization efforts and enzymatic replication efficiency was within 6- to 490-fold lower than a natural base pair.<sup>[9]</sup> Massive PCR amplification of NaM:5SICS containing DNA without significant mutation of the unnatural base pair demonstrated a fidelity of at least 99.9% independent of the sequence context. The error rates of  $10^{-3}$  to  $10^{-4}$  per nucleotide calculated for the unnatural base pair are similar to the  $10^{-4}$  to  $10^{-7}$  observed for natural base pair amplification.<sup>[138]</sup> Despite this achievement, reoptimization of 5SICS **55** resulted in NaM:TPT3 **56:57** as a further improved base pair.<sup>[5]</sup> A PCR-based screen of about 6000 unnatural base pair candidates, confirmed NaM:TPT3 **56:57** as the “best” unnatural base pair known to date, but also found several other pairs, which performed better than the NaM:5SICS pair **56:55**. Some of these alternative candidates are shown in Figure 1-10.<sup>[8]</sup>



**Figure 1-10.** Structural formulas of unnatural base pairs with a better efficiency and fidelity in PCR than the NaM:5SICS base pair **56:55**. R = Sugar and phosphate residues.

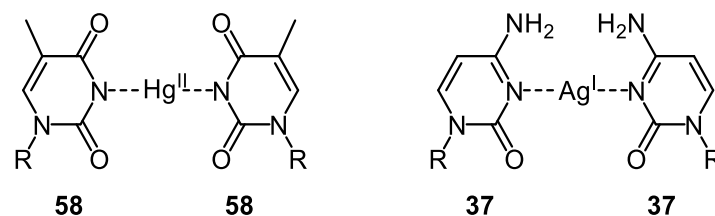
Nevertheless, *in vivo* replication and propagation of the NaM:5SICS unnatural base pair **56:55** was achieved in *E. coli*.<sup>[139]</sup> The artificial nucleotide supply within the host cell was a major obstacle, as intracellular nucleotide biosynthesis from the unnatural nucleosides provided in the medium was insufficient. Finally, a nucleotide transporter was overexpressed in *E. coli*, which allowed uptake of dNaMTP and d5SICS from the medium. A plasmid containing a single NaM:TPT3 base pair **56:57** was



transformed into the cells and was isolated from the culture after 20 h growth. LC-MS/MS analysis of free nucleosides from the digested plasmid detected a single unnatural base pair per plasmid and thus demonstrated the successful *in vivo* replication. In addition, Sanger sequencing and biotin-label based amplification assays were in accordance with mass spectrometry data.<sup>[139]</sup>

### 1.5.4 Metal-Base Pairs

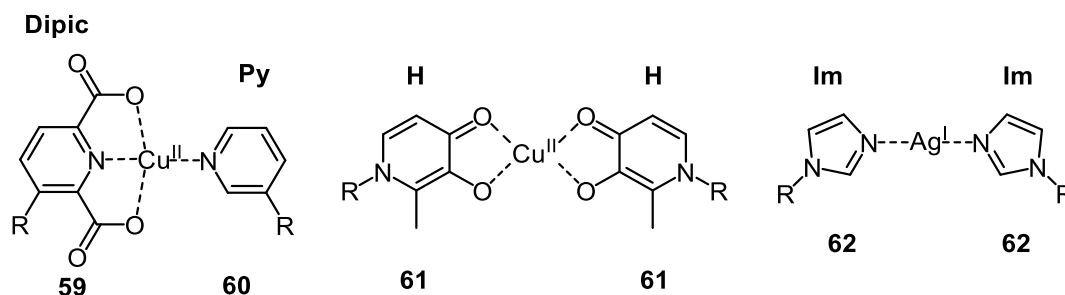
Metal-base pairs represent the third class of unnatural base pairs, which have been studied. Such base pairs incorporate a central metal cation serving as a ligand for coordination. Viscosity changes in thymidine-rich DNA-containing samples in the presence of Hg(II)-ions were already observed in 1952, which prompted the proposal of a thymidine-mercury-thymidine complex (**58** Figure 1-11) by *Katz*.<sup>[140]</sup> More than 50 years later, *Ono* and coworkers showed that a thymidine-thymidine mismatch was thermodynamically stabilized by Hg(II)-ions<sup>[141]</sup> and proved the metal-base pair structure by NMR spectroscopy.<sup>[142]</sup> In the meantime, several natural mismatches have been discovered to serve as a scaffold for metal-cations. For example, C-C (**37**, Figure 1-11),<sup>[143]</sup> C-T,<sup>[144]</sup> and C-A<sup>[145]</sup> mismatches are stabilized by Ag(I)-ions. These findings have led to several DNA-based applications.<sup>[146-147]</sup> For example, selective binding of mercury ions by T-T mismatches has been used to construct a DNA-based Hg(II) sensor.<sup>[146]</sup> Remarkably, the Ag(I)-mediated mismatch formation between C-T or C-C and the T-Hg(II)-T base were selectively formed and elongated by KF or *Thermus aquaticus* polymerase, the latter just for elongation.<sup>[148-149]</sup>



**Figure 1-11.** Structural formula of the T-Hg(II)-T and the C-Ag(I)-C base pair. R = Sugar and phosphate residues.

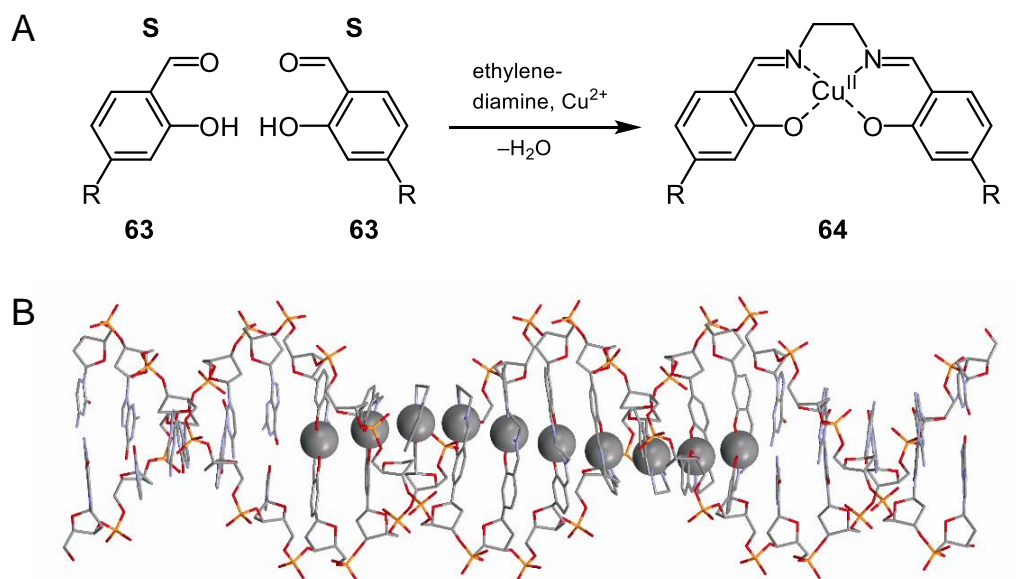
The first metal-base pair between two artificial nucleobases inside DNA was reported by *Meggers et al.* in 2000.<sup>[150]</sup> A DNA duplex containing a central Dipic:Py pair (**59:60**, Figure 1-12) was thermally stabilized in the presence of Cu(II) ions. Therefore, it was suggested that metal-ligand coordinative bonds can potentially replace the hydrogen-bonding for interstrand interaction in unnatural base pairs.<sup>[150-151]</sup> *Shionoya* and coworkers developed a hydroxypyridone variant **H 61**, which forms H-Cu(II)-H pairs (Figure 1-12) and was used to stack several metal-ions inside the DNA helix. Despite these major changes compared to natural DNA double strands, CD-spectroscopy indicated B-form DNA for a duplex containing five consecutive metal-base pairs.<sup>[152]</sup> In the presence of Fe(III) even triple-helix complexes were reported for the H ligand **61** in adequately designed DNA strands.<sup>[153]</sup> Some structural insight was provided by *Johannsen et al.* for a silver-ion mediated imidazole self-pair (**62**, Figure 1-12). A DNA duplex containing

three consecutive Im-Ag(I)-Im metal-base pairs was analysed by NMR spectroscopy. The structure revealed only minor structural changes compared to normal B-form DNA.<sup>[154]</sup>



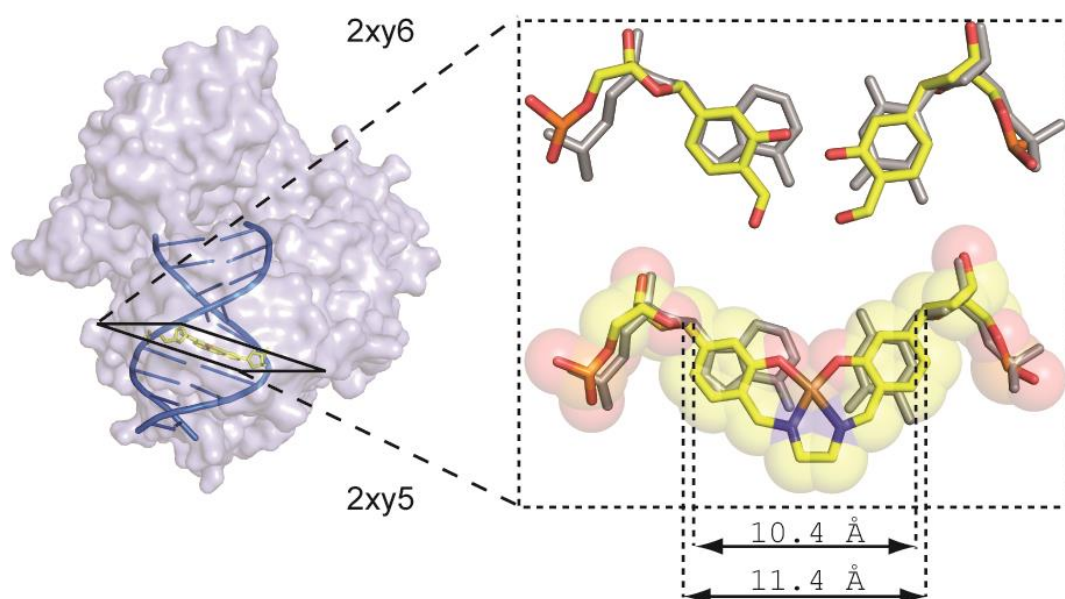
**Figure 1-12.** Selected metal-base pairs, which form inside DNA double strands and involve artificial nucleobases. R = Sugar and phosphate residues.

Inspired by the famous salen ligand from homogenous catalysis,<sup>[155]</sup> *Clever et al.* developed a salicylaldehyde nucleoside **63**. Imine formation upon addition of ethylenediamine crosslinks two opposing salicylaldehydes and the resulting scaffold forms a Cu(II) salen complex **64** (Figure 1-13, A). Due to the properties of the imine bond and the metal complexation, this reaction is reversible in aqueous solutions. In the presence of ethylenediamine and Cu<sup>2+</sup>, DNA double strands containing a single central salicylaldehyde pair displayed a 30 K increase in the melting temperature in comparison to a natural duplex. Compared to other published metal-base pairs this was the strongest thermal duplex stabilization reported at that time.<sup>[12]</sup> Moreover, DNA salen metal complexes for Mn(III), Fe(III), Ni(II) and VO(II) were observed by electrospray ionization mass spectrometry<sup>[156]</sup> and a full helical turn of metal ions, i. e. ten consecutive base pairs, was stacked inside DNA (Figure 1-13, B).<sup>[157]</sup> More recently, electron transfer through a single DNA double-strand containing the salen metal-base pair was studied and displayed an increased conductivity in the presence of the Cu(II) complex.<sup>[158]</sup>



**Figure 1-13.** (A) Two salicylaldehyde bases **63** form the copper salen base pair **64** in the presence of ethylenediamine and a Cu(II)-salt. R = Sugar and phosphate residues. (B) Computer model of the duplex containing ten consecutive Mn(III)-salen complexes from *Clever et al.*<sup>[157]</sup>

Most remarkably, the Cu(II) complexing salen self-pair **64** is capable of being amplified by PCR. In addition, even a X-ray crystal structure was obtained of a DNA polymerase with the salen complex-containing DNA.<sup>[13]</sup> The structure proved that the salicylaldehyde bases were covalently cross-linked by ethylenediamine with a copper ion sequestered in the centre. In comparison to a natural base pair, the Cu(II)-salen duplex was widened (Figure 1-14). Since replication of the salen base pair **64** was faithfully possible in the presence of Cu(II)-ions and ethylenediamine, the next step aimed for is transcription and translation.



**Figure 1-14.**<sup>[13]</sup> Crystal structure of the binary complex between *Geobacillus stearothermophilus* polymerase I and a DNA with a central salen base pair (yellow), with ethylenediamine and copper (bottom) and without (top).

## 2 Aim of the Project

Many modified nucleosides have been introduced into DNA and RNA by the solid-phase synthesis technique using phosphoramidite chemistry. However, this method is limited to provide oligonucleotides of up to 150-200 bases length at its best.<sup>[159]</sup> Longer DNA strands can be obtained by biochemical methods, like the polymerase chain reaction<sup>[160]</sup> based assembly of numerous synthesized oligonucleotides.<sup>[161]</sup> The enzyme machinery involved in handling of oligonucleotides was selected for hundreds of million years to minimize mutation to an acceptable rate.<sup>[162]</sup> As a result, the efficient enzymatic incorporation of non-canonical nucleotides represents a tremendous challenge.

To date a variety of artificial base pairs were described which are independent from hydrogen bonds for base pair formation and rely on hydrophobic interactions.<sup>[8, 11, 138-139, 163]</sup> Notably, metal-complexation has been used as an alternative strategy for unnatural base pair formation. Several nucleobase-metal complexes have been reported inside DNA.<sup>[12, 150-152, 154]</sup> Very few of these so-called metal-base pairs can be formed or extended by polymerases.<sup>[13, 145, 149]</sup> The salen metal-base pair was developed by *Clever et al.* and provides DNA with a highly increased thermal stability.<sup>[12]</sup> Recently, faithful replication of the salen base pair in polymerase chain reactions (PCR) was demonstrated in the presence of  $\text{Cu}^{2+}$  and ethylenediamine.<sup>[13]</sup> Next to the incorporation in PCR, the transcription of artificial base pairs is required for the expansion of the genetic code. Whether a metal-base pair can be transcribed into RNA and even translated into proteins was so far unknown. This research project focussed at the study of transcription and translation of the salen-metal base pair. Therefore, generation of RNA strands with the salicylaldehyde base was to be studied using T7 RNA polymerase as well as solid-phase synthesis. In order to achieve this very challenging goal mutation of the corresponding enzymes was to be analyzed as well.

One drawback of the salen metal-base pair is its dependence on metal complexation for bond formation, which greatly complicates *in vivo* applications. Thus, a novel base pair, independent of metal complexation and external amines, was of interest. A novel base pair was to be developed in allusion to the sterical properties of the salen base pair. Therefore, a set of different amine nucleosides was to be designed, synthesized and analyzed for reversible base pair formation with the salicylaldehyde nucleoside. The effect of such an amine-aldehyde base pair on the DNA duplex stability and structure was to be studied by temperature-dependent UV- and CD-spectroscopy. Candidate base pairs, which stabilize the double strand were to be used for primer extension and X-ray structural analysis.

Fluorophore-modified nucleotides are an important part of modern sequencing methods, since they allow facile imaging of the enzyme reaction.<sup>[103]</sup> Depending on the sequencing method, fluorescent dyes are either introduced to the nucleobase<sup>[164]</sup> or the terminal phosphate.<sup>[3]</sup> Modifications at the terminal phosphate are cleaved upon nucleotide incorporation and thus allow real-time detection of the sequencing reaction.<sup>[7, 165]</sup> Synthesis of phosphate-modified nucleotides can be cumbersome.<sup>[7]</sup> Therefore, another goal of the thesis was the development of a modular strategy relying on Cu(I)-catalyzed azide-alkyne cycloadditions

---

(CuAAC)<sup>[71-72]</sup>. Introduction of short alkyne chains to the  $\gamma$ -phosphate of nucleoside triphosphates, was supposed to allow facile modification with different azido-fluorophores using CuAACs. The resulting labeled nucleotides could be substrates in real-time experiments involving polymerases, like e.g. next-generation sequencing. Therefore, enzymatic acceptance of the labeled nucleotides by DNA, RNA polymerases and kinases had to be studied. In addition, chemoenzymatic introduction of bioorthogonal labels to oligonucleotides was attempted.

### 3 Results and Discussion

In course of this PhD thesis, three major projects were investigated, centered around the chemical synthesis of artificial nucleotides and their application in enzymatic reactions. From the chemical point of view, the synthesis of *C*-nucleosidation was challenging, as there is no established general method and the yields are often lower in comparison to the corresponding *N*-nucleosides. Also, the chemical formation of triphosphates is cumbersome, due to regio- and chemoselectivity issues. Since the subsequent biochemical assays require only small amounts of the nucleotides and oligonucleotides, low-yielding steps in the synthesis were tolerable to a certain extent and tedious optimization was neglected. From the biological side, the specificity of the enzymes for their natural substrates was the greatest obstacle we needed to overcome with respect to the design of the artificial nucleotides. Many DNA and RNA polymerases were optimized during millions of years of evolution to faithfully replicate the oligonucleotide template. Therefore, variation of the *in vitro* conditions during enzymatic assays was crucial, to force incorporation of the artificial substrates. In addition, the consequences of mutagenesis on substrate selection and processivity were investigated in the case of the T7 RNA polymerase. The outcome of the enzymatic reaction was evaluated by polyacrylamide gel electrophoresis and MALDI-TOF mass spectrometry.

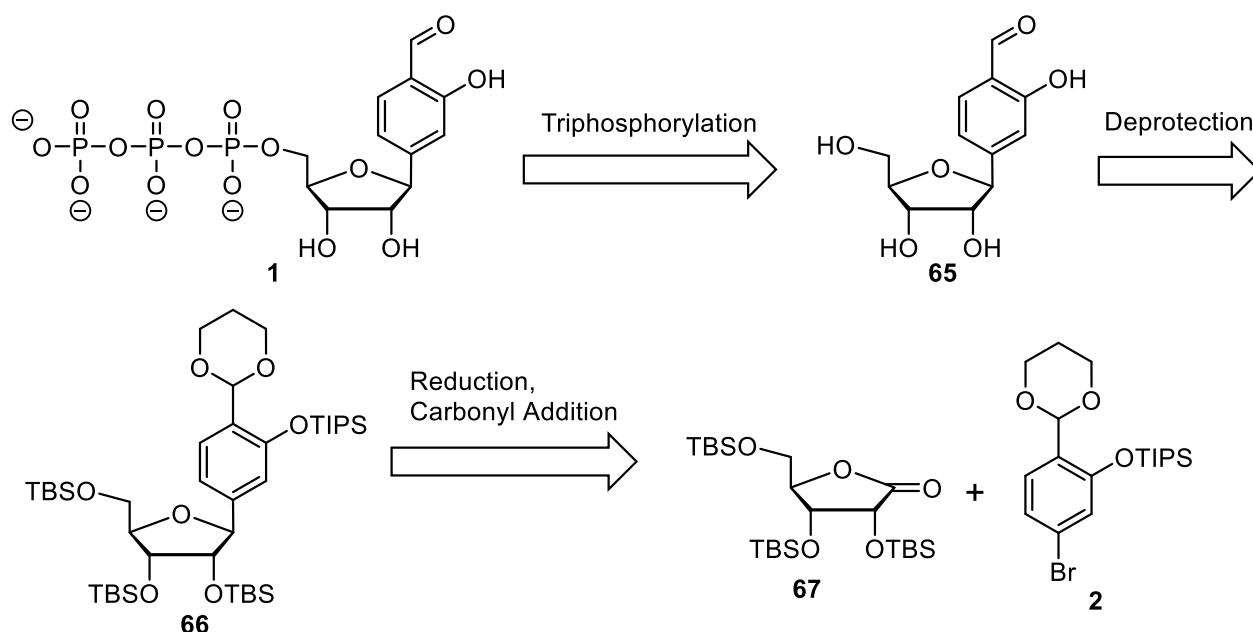
#### Part I – Efforts Towards Faithful Transcription and Translation of a Metal Base Pair

##### 3.1 Synthesis of a Salicylaldehyde Ribonucleotide

The initial focus of this work was to analyze whether the salicylaldehyde base (**S**) pair is capable of being transcribed into RNA. Therefore, the phosphoramidite of the deoxyribose salicylaldehyde was synthesized according to published methods and used to generate the DNA template.<sup>[12]</sup> In addition, a ribonucleotide form of the salicylaldehyde base was needed for the incorporation by the RNA polymerase. Moreover, a salicylaldehyde RNA phosphoramidite was to be synthesized to allow solid-phase synthesis of RNA strands that could be used as authentic standards for the enzymatic experiments.

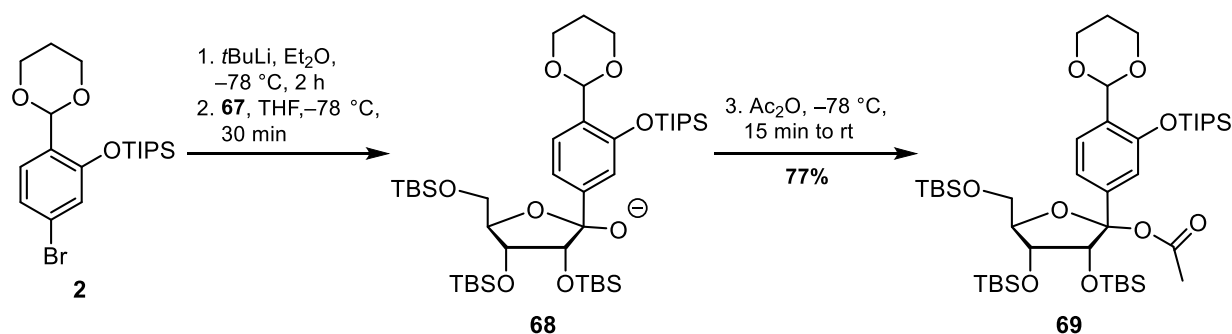
The synthesis of the salicylaldehyde triphosphate **1** (STP) was designed to have a final triphosphorylation step of a salicylaldehyde nucleoside **65** (

Scheme 3-1), which could be accessed from its protected nucleoside **66** by silyl ether and acetal cleavage. In a convergent reaction, this nucleoside **66** was synthesized by nucleophilic addition of a protected bromo salicylaldehyde **2** to the carbonyl of a TBS-protected ribolactone **67** and subsequent reduction following a method described by Štefko *et al.*<sup>[39]</sup> The TBS-ribolactone **67** could be obtained in a two-step one-pot sequence from natural D-ribose according to published protocols.<sup>[166-167]</sup> To provide multigram amounts of the protected bromo salicylaldehyde **2**, which was also needed for the synthesis of the corresponding DNA compound, a three-step procedure starting with 3-bromophenol was performed as described by Clever *et al.*<sup>[12]</sup>



**Scheme 3-1.** Retrosynthetic analysis of salicylaldehyde triphosphate **1** (STP).

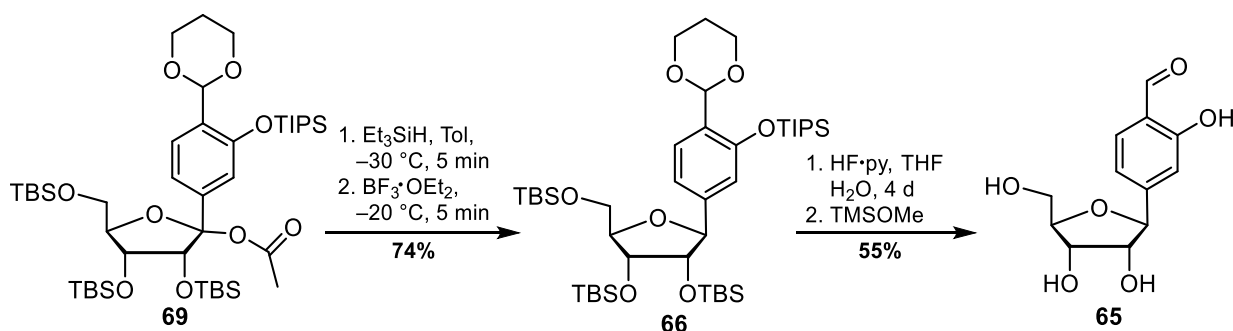
The key step of the salicylaldehyde triphosphate **1** synthesis was the *C*-glycosylation reaction. After a bromide-lithium exchange at the protected salicylaldehyde **2** using *t*BuLi at  $-78\text{ }^{\circ}\text{C}$ , the TBS-ribolactone **67** was added and the resulting deprotonated hemiketal **68** was trapped with acetic anhydride (Scheme 3-2). Despite the sensitivity of the acetylated nucleoside **69** towards heat and acid, modifications of the initial protocol enabled synthesis of the nucleoside precursor **69** with 77% isolated yield. In contrast to this, the non-acetylated hemiketal **68** was chemically stable under neutral and slightly acidic conditions and could be isolated in 84% yield, when the acetylation step was omitted. Based on NMR data, both hemiketals were exclusively isolated as the  $\beta$  anomer, as reported for the synthesis of related *C*-ribonucleosides.<sup>[39]</sup>



**Scheme 3-2.** Synthesis of acetylated nucleoside **69** from protected salicylaldehyde **2** and TBS-protected ribolactone **67**.

In order to eliminate the additional functional group and to obtain the nucleoside **66**, a subsequent reduction using a hydride source under Lewis acid conditions was necessary. Therefore, the procedure from Štefko *et al.* was adapted with regards to temperature and solvent. The reaction starts with a

deoxygenation to an oxonium intermediate that is subsequently reduced by the hydride source. In the final protocol  $\text{Et}_3\text{SiH}$  and  $\text{BF}_3 \cdot \text{OEt}_2$  in toluene at  $-20^\circ\text{C}$  provided the desired protected  $\beta$ -nucleoside **66** in 74% yield (Scheme 3-3). The quality of the reagents and a strict temperature control was crucial for the success of the reaction. Moreover, only the chemically unstable acetylated hemiketal **69** was reasonably converted under the reaction conditions, whereas no conditions were found to provide more than traces of nucleoside **66** from the reduction of the non-acetylated hemiketal **68**. Instead, Lewis acid-mediated deprotections of the acetal and silyl ether groups were observed. This illustrates the role of the leaving group on the speed of the oxonium intermediate formation. Also, no  $\alpha$  anomer of the protected nucleoside **66** could be isolated when using toluene as a solvent.

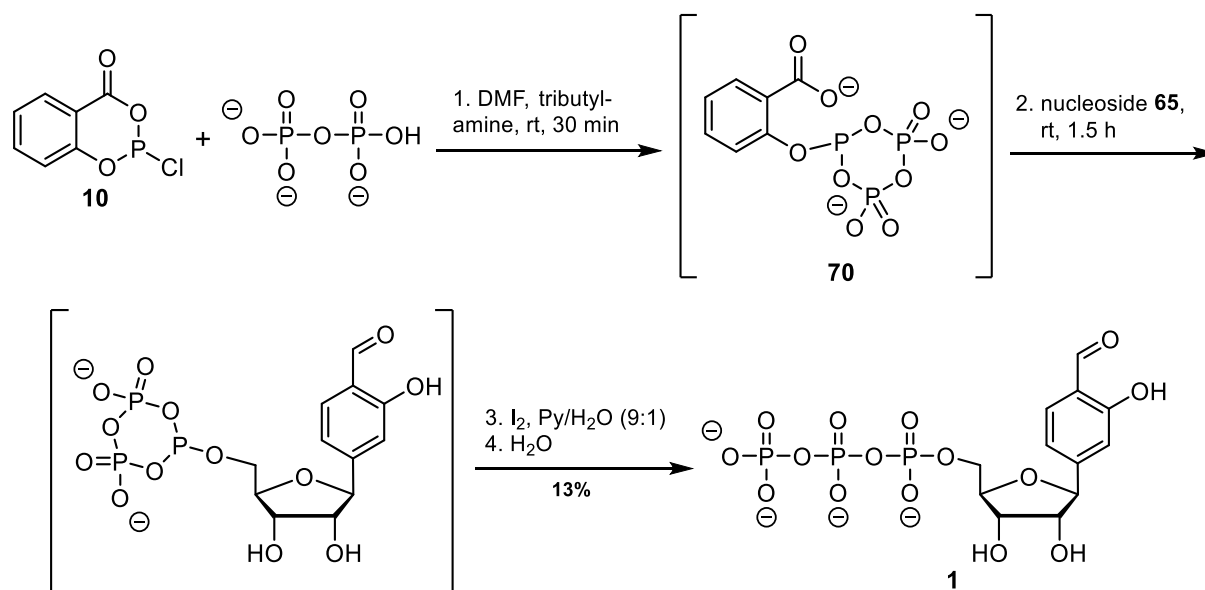


**Scheme 3-3.** Synthesis of the salicylaldehyde nucleoside **65** from acetylated hemiketal **69**.

For the subsequent global deprotection to the salicylaldehyde nucleoside **65**  $\text{HF} \cdot \text{pyridine}$  was used in a  $\text{THF}/\text{H}_2\text{O}$  mixture. This enabled the cleavage of all silyl ether groups and hydrolysis of the acetal with a moderate isolated yield of 55%. Unfortunately, flash column chromatography and the work-up procedure could not separate the 1,3-propanediol released from the acetal deprotection completely. After HPLC-purification the desired nucleoside **65** was pure enough to be used in the triphosphate reaction setups.

In contrast to the enzyme-catalyzed biosynthesis of triphosphates, the chemical synthesis of these charged and highly functionalized molecules can be cumbersome. Even famous methods like the one by *Ludwig* and *Ruth* from the early eighties (Scheme 1-5, p. 6),<sup>[48-49]</sup> which represented the only feasible protocol for many years that was widely used, proved to be incompatible with the synthesized artificial nucleotides and resulted in only minor triphosphate yields. The application of this method to the salicylaldehyde deoxynucleoside provided only 2% yield of the corresponding triphosphate after several rounds of RP-HPLC purification.<sup>[13]</sup> In 2011 a further improved protocol of the *Ludwig* and *Ruth* method was published by *J. Caton-Williams et al.*,<sup>[51]</sup> which relied on a salicylic phosphite reagent **10** instead of  $\text{POCl}_3$  (Scheme 3-4). This phosphite **10** reacts with tributylammonium pyrophosphate to generate a “triphosphate reagent” **70** *in situ* (Scheme 3-4), which is less reactive and displays a better regioselectivity when applied to a nucleoside. Subsequent oxidation and hydrolysis is followed by  $\text{NaCl}$ -ethanol precipitation, which yields crude triphosphates of sufficient purity for being accepted by polymerases.<sup>[51, 64]</sup>

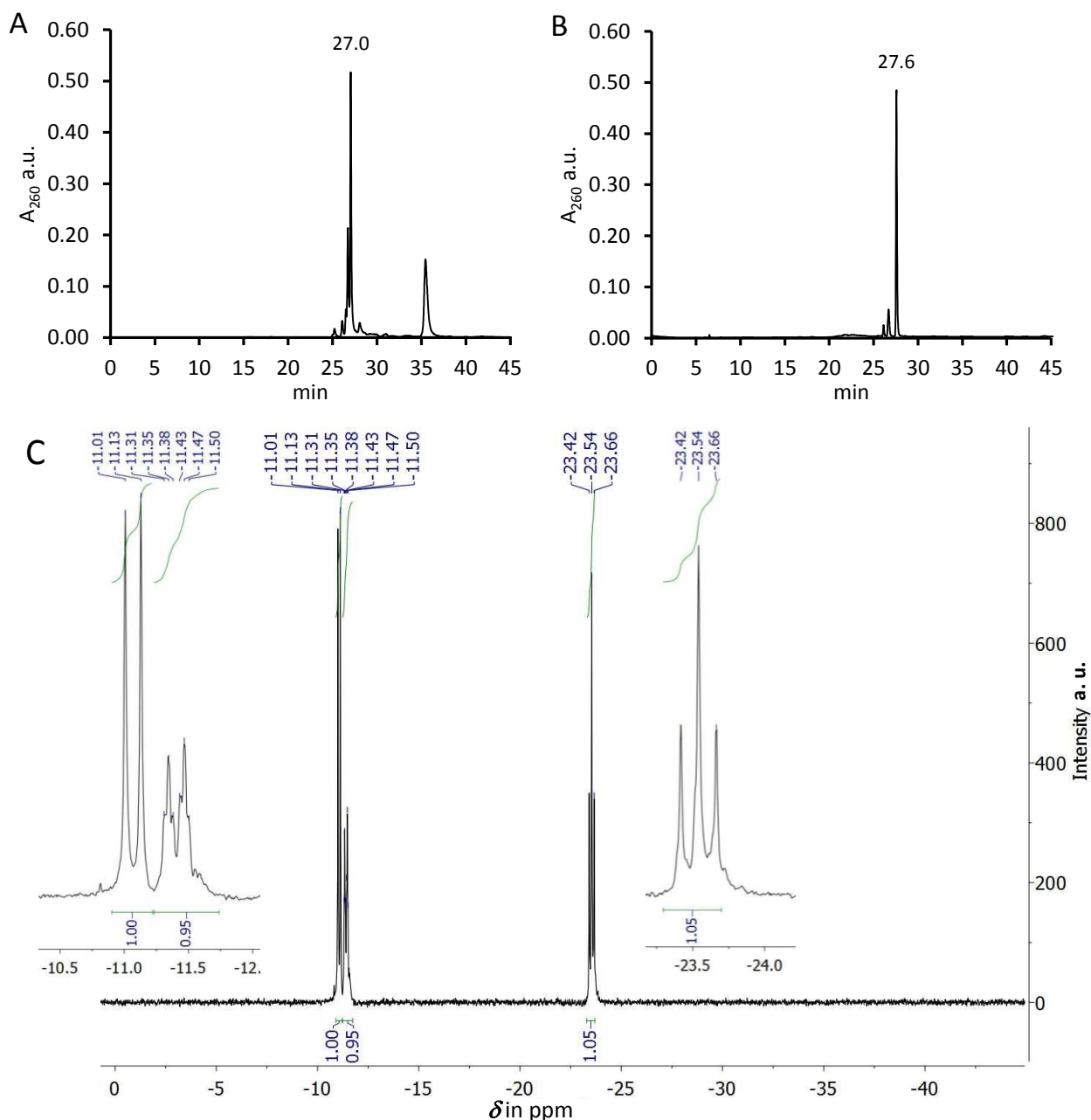




**Scheme 3-4.** Suggested mechanism of the salicylic ribonucleotide **1** synthesis using a salicylic phosphite reagent **10** adapted according to *J. Caton-Williams et al.*<sup>[51]</sup>

When this novel protocol was applied to the salicylaldehyde ribonucleoside **65** an isolated yield of 13% was obtained for the ribonucleotide **1**, which significantly improved the yield compared to the corresponding deoxyribonucleotide. To avoid inhibitory contaminations for the RNA polymerase in the transcription assay, the crude ribonucleotide **1** was purified twice by RP-HPLC. The resulting salicylaldehyde triphosphate (STP) **1** was analyzed by analytical RP-HPLC, high-resolution ESI mass spectrometry and NMR. It is noteworthy that the crude STP **1** which was obtained after NaCl-ethanol precipitation of the reaction was already reasonably pure, as the main peak corresponds to the isolated triphosphate **1** (Figure 3-1, A and B). Even after two-fold purification, minor peaks were observed in the chromatogram in addition to the strong signal at 27.6 min (Figure 3-1, B). ESI MS revealed that all peaks possessed the STP molecular mass. In order to exclude contamination by triphosphate (regio)-isomers, the NMR spectra were carefully analyzed. The <sup>1</sup>H-decoupled <sup>31</sup>P NMR spectrum showed three signals, from the triphosphate phosphorus atoms. Due to the <sup>2</sup>J<sub>P-P</sub> coupling two doublets and one triplet were observed from -23 to -11 ppm (data not shown). The triplet was readily assigned to the β-phosphate phosphorus. In order to assign the remaining signals, a proton-coupled <sup>31</sup>P NMR was measured. The <sup>3</sup>J<sub>P-H</sub> coupling between the α-phosphate phosphorus and a nucleoside proton changed the doublet at -11.5 ppm to a doublet of triplets (Figure 3-1, C). This indicates a coupling to a methylene proton. Since the salicylaldehyde nucleoside only possesses a single methylene group at the 5'-position, this finding confirms that the main product is the desired 5'-STP. The remaining doublet had to originate from the γ-phosphate. All signals of the <sup>1</sup>H NMR were readily assigned aided by COSY and HSQC spectra. No signals except those of the salicylaldehyde 5'-triphosphate were found in <sup>1</sup>H NMR and <sup>31</sup>P NMR spectra. This means that the purified STP was at least 95% pure despite the minor peaks detected by analytical RP-

HPLC. Probably, depending on the number of counterions of the 5'-triphosphate during HPLC, minor shifts of the retention time can occur.

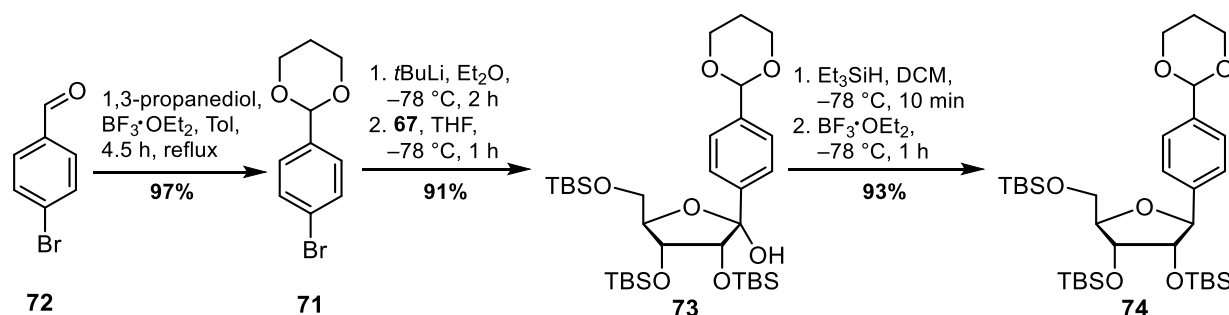


**Figure 3-1.** Analytical HPLC profiles of crude (A) and purified (B) STP 1 at 260 nm, 0-30% B in 45 min. (C)  $^{31}\text{P}$  NMR spectrum of the purified STP. The three signals were assigned to the  $\gamma$ -,  $\alpha$ - and  $\beta$ -phosphate, respectively. The doublet of triplets at  $-11.4$  ppm for the  $\alpha$ -phosphate proves the formation of the 5'-triphosphate.

### 3.2 Synthesis of a Benzaldehyde Ribonucleotide

In order to study the role of the *ortho*-hydroxy group in the salicylaldehyde base, a benzaldehyde nucleoside was synthesized, transformed to the triphosphate and then also applied in the transcription assay. The synthesis of the benzaldehyde nucleoside was achieved based on the established route for the salicylaldehyde nucleoside. This time, the benzaldehyde base building block **71** was readily available in

one step from commercial 4-bromobenzaldehyde (**72**) in 97% yield, without the need for flash column chromatography.<sup>[168]</sup> The subsequent *C*-glycosylation using the protocol from the salicylaldehyde nucleoside synthesis without acetylation, yielded **73** in an excellent 91% after isolation, even on a multigram scale (Scheme 3-5). However, when the *in situ* acetylation was applied, the resulting acetylated hemiketal was not stable enough to withstand flash column chromatography. Instead, the crude product containing about 60% (judged by <sup>1</sup>H NMR) of the acetylated product was used directly in the subsequent reduction to the protected nucleoside **74**. Under the conditions that were successful for the synthesis of the salicylaldehyde nucleoside, no desired protected nucleoside **74** was obtained. In contrast, however the non-acetylated hemiketal could be reduced to the protected nucleoside **74** in an excellent 93% isolated yield, using a slightly modified protocol. Moreover, in addition to the fully protected nucleoside **74**, 4% of TBS-protected nucleoside without the acetal group was detected, which formed from Lewis acid-promoted deprotection during the reduction. As the subsequent reactions to the benzaldehyde nucleoside were deprotection steps, this side product was valuable. Interestingly, the minor change of the base building block for the benzaldehyde nucleoside compared to the salicylaldehyde had a pronounced effect on the reactivity.

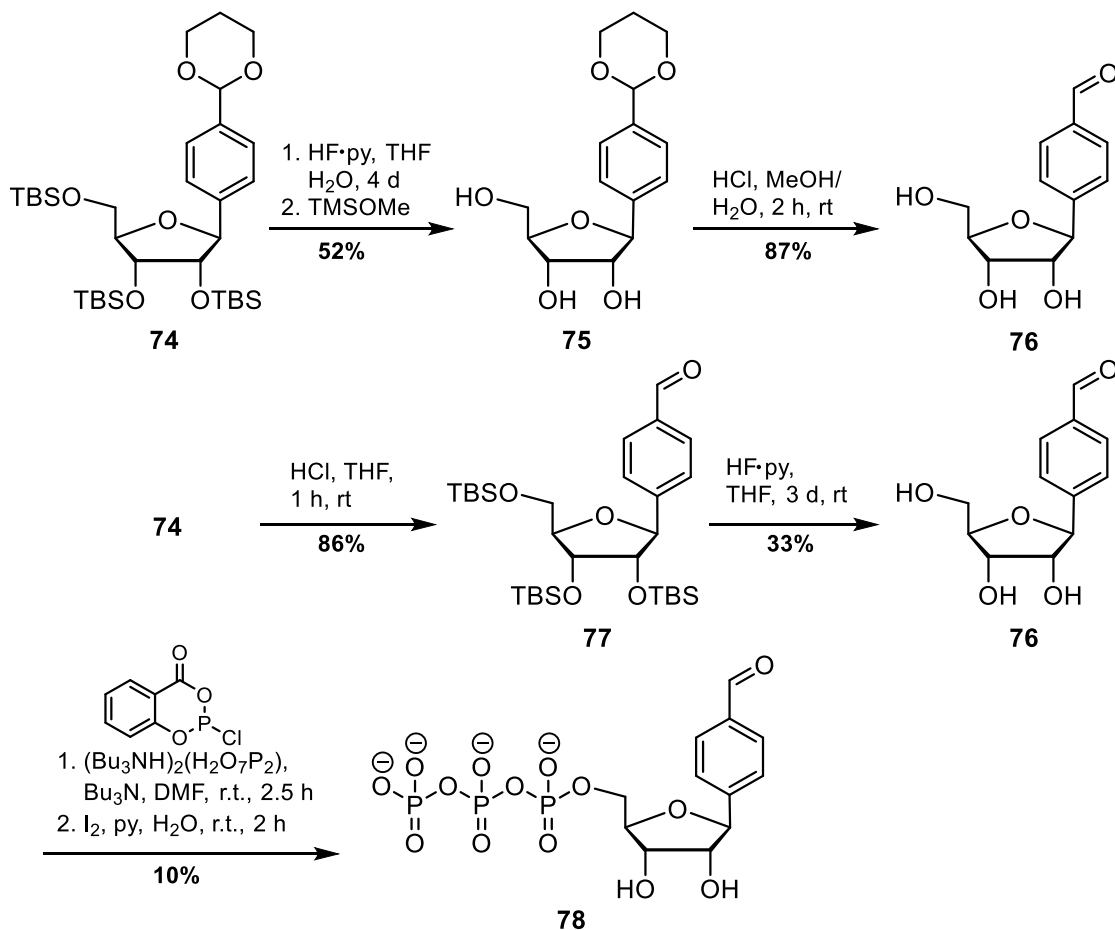


**Scheme 3-5.** Synthesis of the protected benzaldehyde nucleoside **74** starting from commercially available 4-bromobenzaldehyde (**72**).

For the global deprotection of the benzaldehyde nucleoside **74**, two strategies were tested. At the beginning, a combined silyl ether and acetal deprotection approach according to the known procedure for the similar salicylaldehyde nucleoside synthesis was examined. However, this resulted exclusively in silyl ether cleavage, giving 52% of the acetal nucleoside **75** (Scheme 3-6). Subsequent acetal cleavage using HCl in a MeOH/H<sub>2</sub>O mixture provided the benzaldehyde nucleoside **76** in good 87% yield, but unfortunately the 1,3-propanediol was only separable by RP-HPLC.

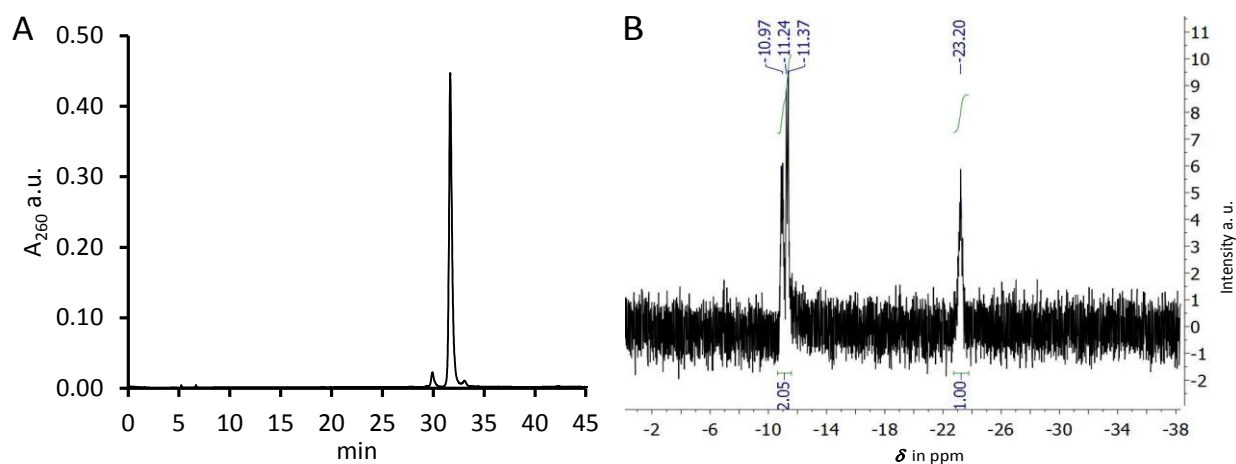
In order to avoid the time-consuming HPLC purification, in the second deprotection strategy the reaction steps were simply reversed. The acidic acetal deprotection using an aqueous HCl/THF mixture gave the TBS-protected benzaldehyde nucleoside **77** in 86% isolated yield (Scheme 3-6). As expected, standard flash column chromatography was sufficient to separate the polar 1,3-propanediol from the much less polar TBS-protected benzaldehyde nucleoside **77**. Silyl ether cleavage using HF·pyridine provided only 33% of the benzaldehyde nucleoside **76** and in total 52% of partly deprotected side products. The one- and

twofold-deprotected products were subjected to another round of silyl deprotection. Efforts replacing HF as a fluoride source by TBAF suffered from purification difficulties. As the obtained amounts of benzaldehyde nucleoside **76** were sufficient for the triphosphate reaction, further optimization of the silyl deprotection reaction was not performed.



**Scheme 3-6.** Synthesis of the benzaldehyde nucleoside triphosphate **78** from protected nucleoside **74** by acetal hydrolysis, silyl ether deprotection and triphosphate formation.

The reaction to the desired benzaldehyde nucleoside triphosphate **78** was performed according to the protocol from *J. Caton-Williams et al.*<sup>[51, 64]</sup> Fortunately, the first attempt already resulted in successful 5'-triphosphate formation and after a two-fold RP-HPLC purification the benzaldehyde ribonucleotide **78** was isolated in 10% yield. Based on analytical RP-HPLC the purity was estimated to be about 95% (Figure 3-2, A). Due to a low signal to noise ratio of the proton-coupled <sup>31</sup>P NMR spectrum (Figure 3-2, B), an additional <sup>31</sup>P-<sup>1</sup>H HMBC measurement was necessary to unambiguously prove the identity of the 5'-triphosphate (data not shown).

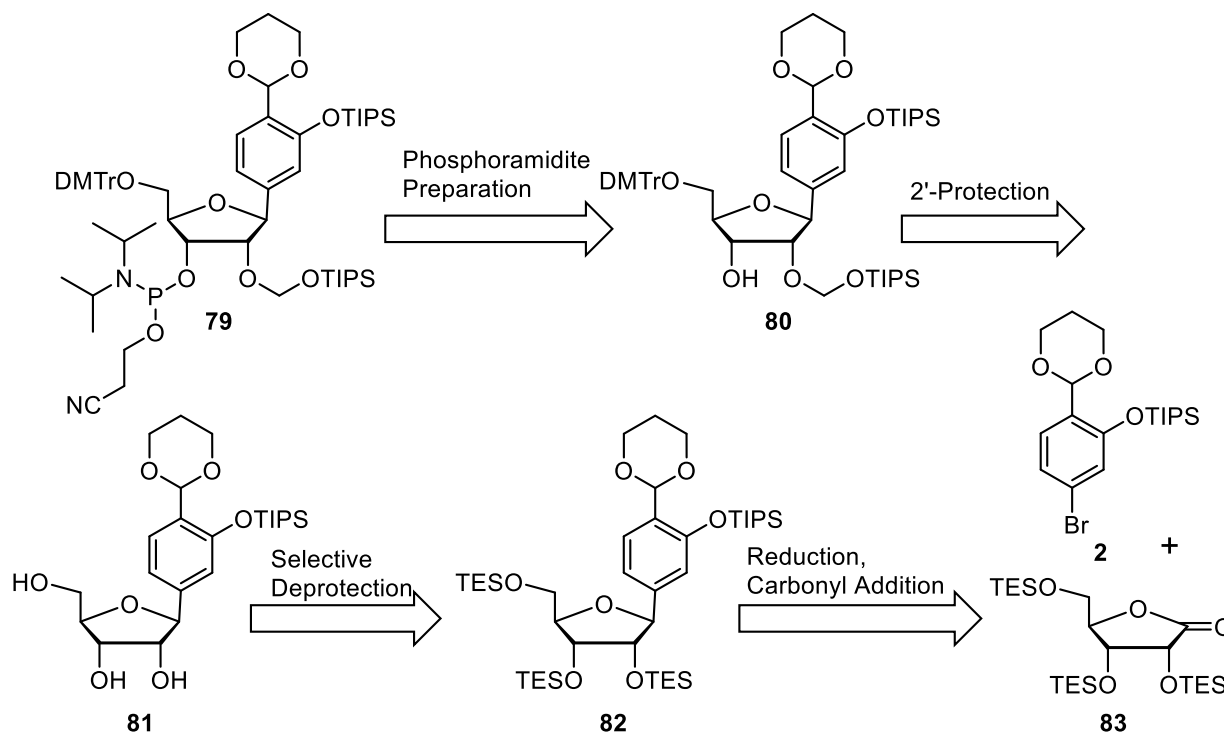


**Figure 3-2.** (A) Analytical RP-HPLC trace and (B)  $^{31}\text{P}$  NMR spectrum of the purified AldTP **78**. HPLC conditions: 0–20% B in 45 min, detection at 260 nm.

In conclusion, the benzaldehyde nucleoside triphosphate **78** synthesis was achieved in a six-step sequence from commercial starting materials. Although the benzaldehyde nucleoside just lacks the *ortho*-hydroxyl group of the salicylaldehyde, considerable differences were observed regarding the reactivity of the molecule in the synthesis. Most notably, the yield of the addition-reduction *C*-glycosylation procedure increased from 56% for the salicylaldehyde to 84% for the benzaldehyde derivative. Also, an enhanced stability of the acetal protecting group was noticed. A neighboring group effect could account for the more facile acetal hydrolysis during silyl deprotection of the salicylaldehyde compared to the benzaldehyde derivative.

### 3.3 Synthesis of a Salicylaldehyde Ribophosphoramidite

In case of a successful salicylaldehyde transcription, it was considered advantageous to have access to RNA strands containing the artificial base for analytical purposes. Therefore, the synthesis of the salicylaldehyde ribophosphoramidite **79** was attempted. This compound in principle allows also the synthesis of salicylaldehyde-containing RNA strands for metal chelation studies in RNA-RNA and RNA-DNA templates.

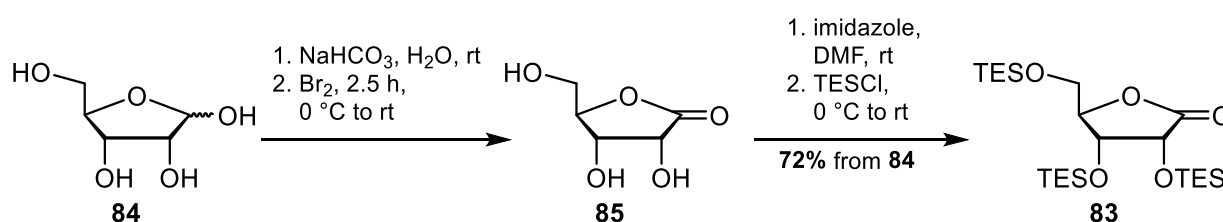


**Scheme 3-7.** Retrosynthetic plan for the synthesis of the salicylaldehyde ribophosphoramidite **79**.

The route towards the ribophosphoramidite **79** was designed based on the synthesis of its triphosphate STP **1** (Scheme 3-2), as the key-step *C*-glycosylation provided the  $\beta$ -configured nucleoside in good yield. The final phosphoramidite was prepared according to standard protocols from the 2'-OTOM-protected nucleoside **80** (Scheme 3-7). According to a published procedure, which was claiming selective 2'-OH protection for a base-protected ribonucleoside, the TOM group was selected.<sup>[169]</sup> Moreover, in comparison to the TBS group, this group enormously decreases the coupling time during solid-phase synthesis. The main difference to the synthesis of the ribonucleotide was the need for an orthogonal hydroxyl protecting group of the sugar moiety and the base building block, to obtain the base-protected ribonucleoside **81**. Triethylsilyl (TES) groups were chosen for the sugar hydroxyl group, as mild acidic cleavage protocols are known, allowing selective deprotection of the TES-protected salicyl ribonucleoside **82**.<sup>[170]</sup> Following the established *C*-glycosylation procedure, the carbonyl addition of the lithiated base building block **2** to a TES-ribolactone **83** and subsequent reduction gave the protected ribonucleoside **82**. The TES ribolactone

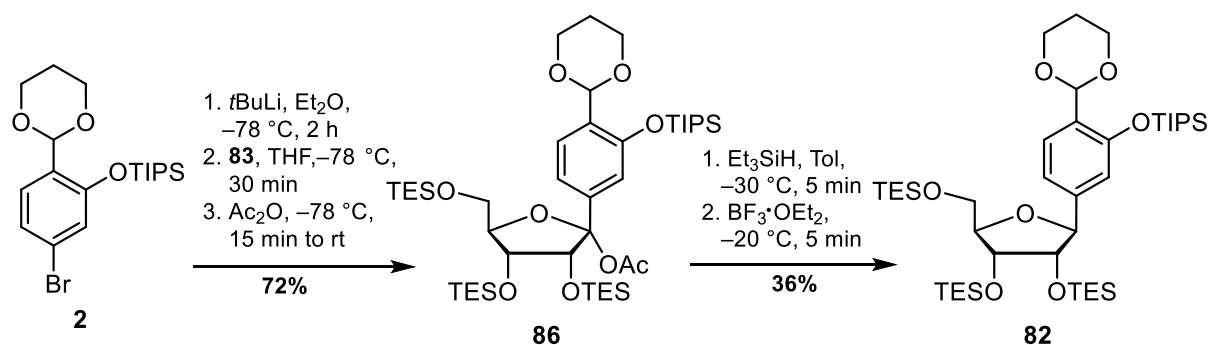
**83** was to be accessed *via* a similar procedure as the TBS variant which had been applied in the synthesis of the STP.

The TES ribolactone **83** was synthesized in a one-pot procedure from D-ribose (**84**) using a combination of published protocols.<sup>[166, 170]</sup> After bromine-mediated oxidation of ribose **84** to lactone **85**, the extensively dried crude product was triethylsilyl-protected on all hydroxyl groups (Scheme 3-8). On a 20 mmol scale this afforded the TES ribolactone **83** in 72% isolated yield. However, when the scale was increased to 40 mmol, the yield dropped to 38%. Therefore, parallel reactions were performed on 10-20 mmol scale to obtain a combined amount of over 20 g TES ribolactone **83** for the subsequent reactions.



**Scheme 3-8.** Synthesis of the TES ribolactone **83** from D-ribose (**84**).

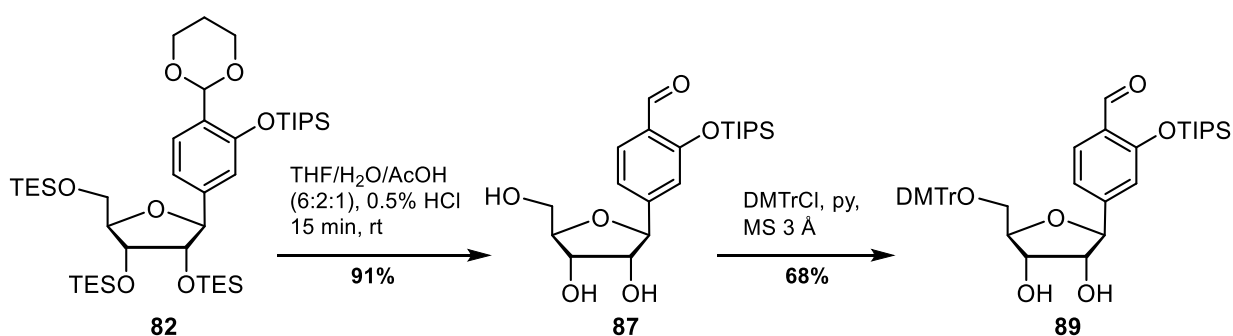
The C-glycosylation reaction was performed according to the conditions established for the STP synthesis (Scheme 3-2). The salicylaldehyde building block **2** was lithiated and added to the TES ribolactone **83** followed by acetylation *in situ*. This yielded the acetylated TES hemiketal **86** after flash column chromatography in 72% (Scheme 3-9), quite comparably to the 77% yield obtained for the TBS analog. In the subsequent reduction to the protected TES nucleoside **82**, only 36% of the desired product was isolated, which was only half of the amount that was obtained for the TBS variant. This was attributed to the decreased acid stability of the TES- compared to the TBS-groups.



**Scheme 3-9.** Synthesis of the TES-protected salicyl ribonucleoside **82** from salicylaldehyde base building block **2** and TES ribolactone **83**.

With the TES-protected nucleoside **82** in hand, several conditions were tested for the selective deprotection of the TES groups. At the beginning, acid-mediated cleavage using NaHSO<sub>4</sub>·SiO<sub>2</sub> as a heterogenous catalyst was studied in DCM.<sup>[171]</sup> Initial test reactions involving the salicylaldehyde base building block **2** and the TES ribolactone **83** were very promising. Whilst the TIPS group remained untouched during 16

hours at 30 °C, the TES ribolactone **83** was completely deprotected under the same conditions. Only minor acetal hydrolysis (<10%) was detected for the base building block **2**. However, when the conditions were applied to the TES-protected nucleoside **82**, mostly decomposition of the compound was observed within a few hours. Next, various THF/AcOH<sub>(aq)</sub> mixtures were studied for the selective deprotection of the TES-groups. A mixture of THF/AcOH/H<sub>2</sub>O (3:1:1) for 13 h at rt for example yielded about 22% of the desired acetal nucleoside **81** and about 87% of the TIPS nucleoside **87** as judged by <sup>1</sup>H NMR. Unfortunately, the yield of the acetal nucleoside **81** could not be improved further. Also, reprotection of the aldehyde in the TIPS nucleoside **87** using a literature-known procedure failed.<sup>[172]</sup> Therefore, it was decided to prepare the aldehyde ribophosphoramidite **88**, since several phosphoramidites are published containing free aldehydes as a functional group.<sup>[173-174]</sup> The deprotection procedure was adjusted accordingly and the TIPS nucleoside **87** was obtained in a very good yield of 91% after isolation (Scheme 3-10).



**Scheme 3-10.** Synthesis of the DMTr-protected TIPS nucleoside **89** from TES-protected nucleoside **82**.

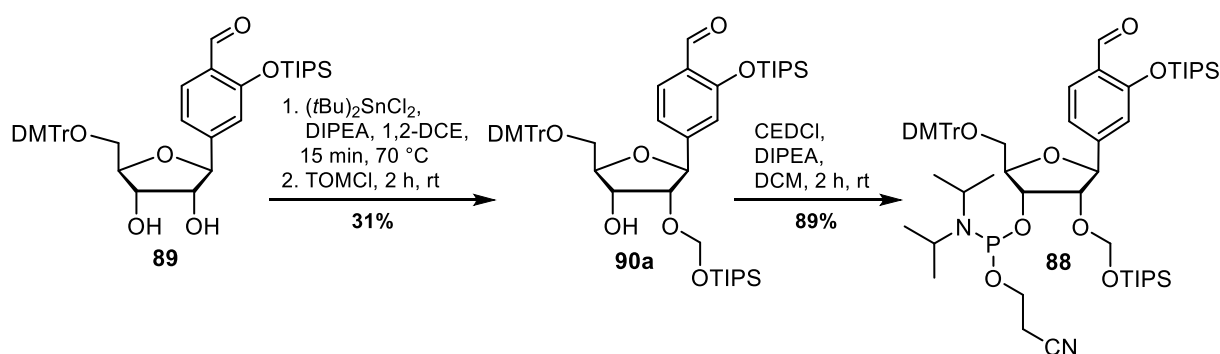
In order to judge the compatibility of the aldehyde group for solid-phase synthesis, the TIPS nucleoside **87** was incubated for three days in the oxidation solution. Satisfyingly, no detectable amount of oxidation or decomposition products were found.

Since partly deprotected products were observed during reduction of the acetylated TES hemiketal **86**, subsequent acidic cleavage was performed using the crude product directly. This afforded the TIPS nucleoside in 61% isolated yield over the two steps, which is almost twice as much as for the single steps combined (32%). It also saved time for flash column chromatography. The DMTr protection was performed using standard conditions on an almost 1 g scale and yielded the 5'-O-DMTr TIPS nucleoside **89** in 68% after isolation.

Next, 2'-OH protection with (triisopropylsiloxy)methyl chloride (TOMCl) was attempted following published procedures, which claim partly selective protection in 40-75% isolated yield for various DMTr-ribonucleosides.<sup>[169,175-177]</sup> Initially, a tin acetal is formed from the hydroxyl groups of the nucleoside in the presence of di-*tert*butyl tin dichloride and DIPEA, followed by addition of TOMCl. The selectivity between 2'- and 3'-protection is described to reach up to 5:1 for this protocol.<sup>[176]</sup> When this method was applied to the 5'-O-DMTr TIPS nucleoside **89**, a 1:1 mixture of 2'- and 3'-TOM-protected nucleosides was obtained with 68% yield. After painstaking repetitive flash column purification the desired 2'-TOM protected nucleoside **90a** was separated from the 3'-isomer in 31% yield (Scheme 3-11). The material had

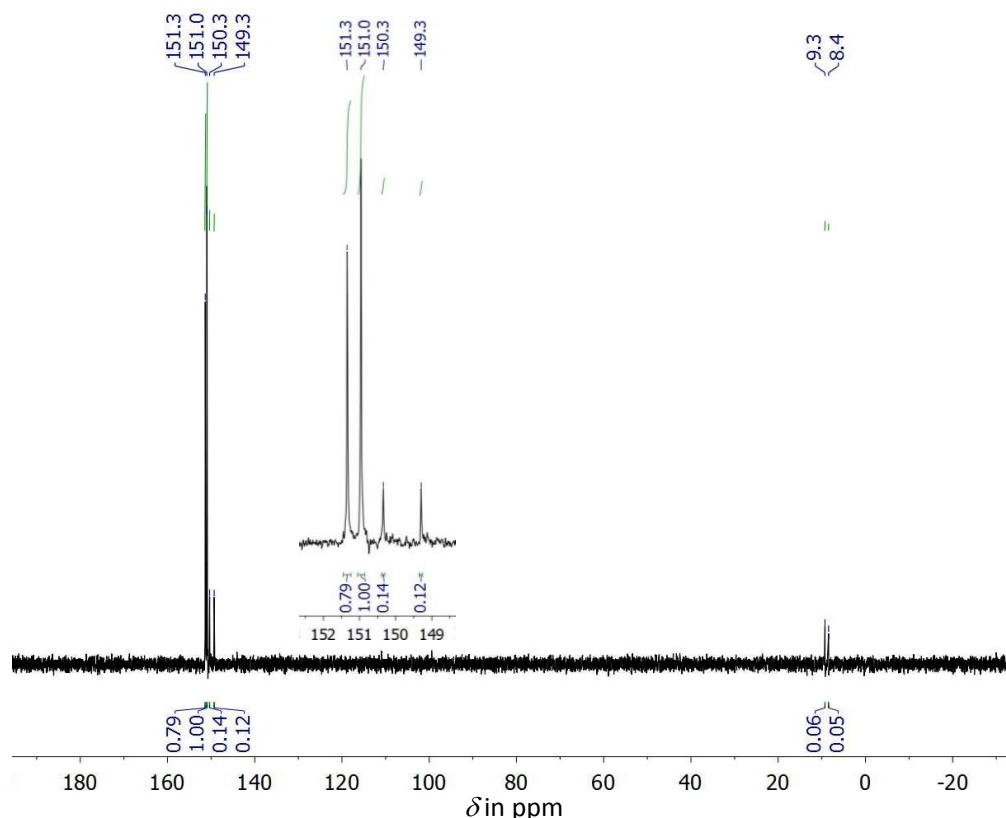


a final purity of  $\geq 90\%$ . Careful analysis of the NMR spectra allowed unambiguous identification of the regioisomers and the complete assignment of all signals. An alternative approach using TBSCl and an excess of imidazole in pyridine just resulted in the formation of the corresponding iminium adduct between the imidazole and the aldehyde of the DMTr nucleoside **89**.<sup>[178]</sup> No TBS protected product was observed. Even when the imidazole was replaced by DIPEA only minor amounts (about 10%) of TBS-protected nucleoside were isolated as a regioisomeric 1:1 mixture. In a final effort to improve the 2'-OH protection issue, a special protocol involving a chiral organocatalyst was applied, which stated a 2':3' selectivity of 98:2 in up to 91% yield.<sup>[179]</sup> This time, selective 2'-O-TBS protection was observed with an acceptable yield of about 50%. However, concomitant loss of the TIPS-group on the nucleobase rendered the resulting compound worthless. All further efforts to improve the reaction were abandoned.



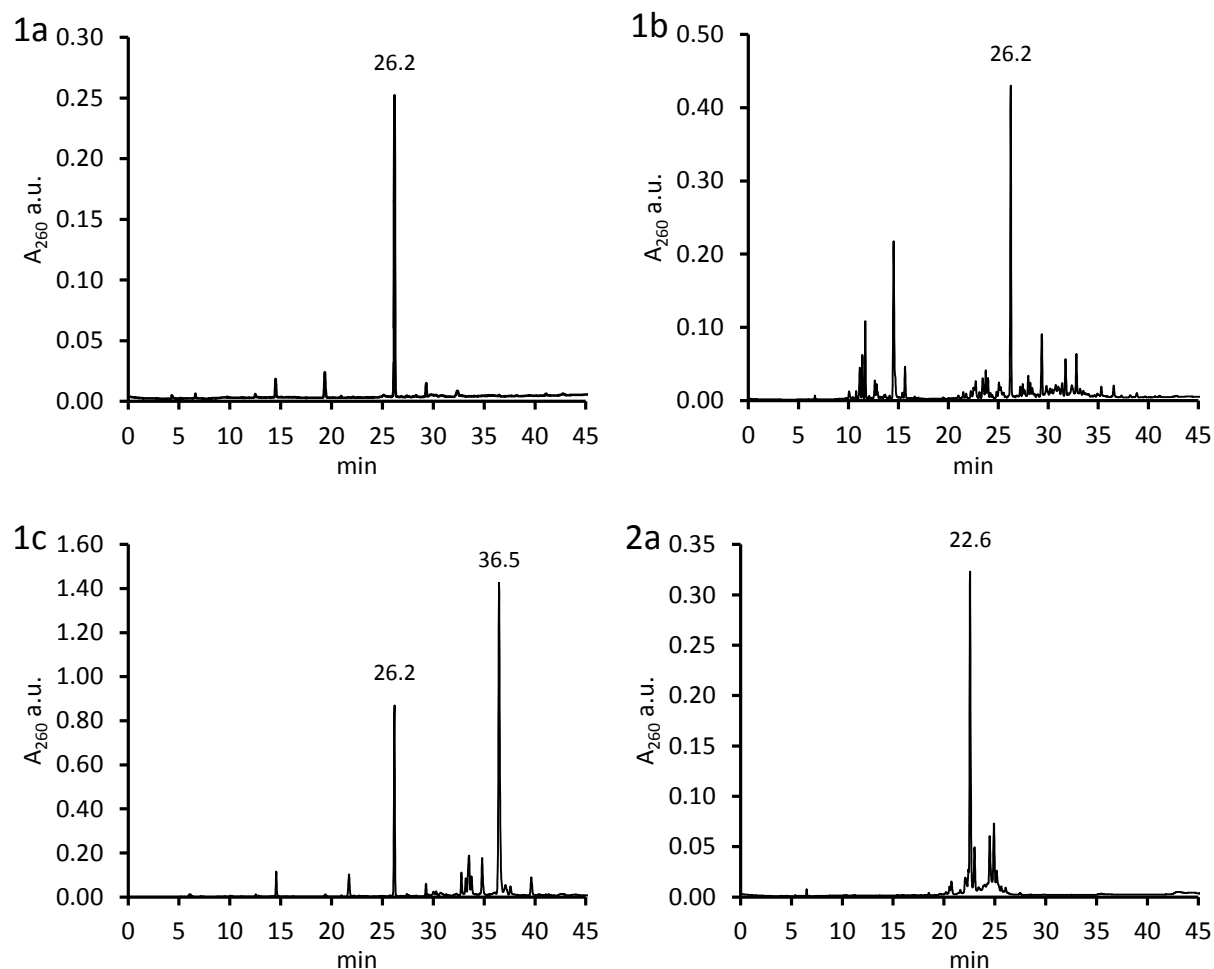
**Scheme 3-11.** Synthesis of the salicylaldehyde ribophosphoramidite **88** from the 5'-O-DMTr TIPS nucleoside **89**.

Then, the obtained 2'-O-TOM-protected DMTr TIPS nucleoside **90a** was converted into the corresponding phosphoramidite. Using standard conditions, the desired salicylaldehyde ribophosphoramidite **88** was obtained in a good yield (89%, Scheme 3-11) and purity as judged by  $^{31}\text{P}$  NMR. The main two signals at 151.3 and 151.0 ppm in the  $^{31}\text{P}$  NMR spectrum are the P-diastereoisomers of the desired phosphoramidite **88**. About 10% impurity was caused by the phosphoramidite product (150.3 and 149.3 ppm) from the 3'-O-TOM nucleoside **90b**, which was already present in the 2'-O-TOM nucleoside **90a** starting material, another 5% was lost as an H-phosphonate (9.3 and 8.4 ppm, Figure 3-3). Without additional purification, the phosphoramidite **88** was directly applied in the solid-phase RNA synthesis.



**Figure 3-3.**  $^{31}\text{P}$  NMR spectrum of the salicylaldehyde ribophosphoramidite **88**.

According to the trityl values, this phosphoramidite **88** had a comparable coupling efficiency as any standard RNA phosphoramidite (data not shown), when it was applied in the solid-phase synthesis of RNA strands. For the deprotection procedure different protocols were studied with a 15mer strand containing a single salicylaldehyde ribonucleoside (5'-CACATTASTGTTGTdA-3'). The initial cleavage was performed in three ways: a) saturated aqueous ammonia in ethanol (1:3) at rt for 18 h, b) 0.4 M NaOH in methanol at rt for 18 h and c) a mixture of saturated aqueous ammonia and methylamine at 65 °C for 10 min. Analytical RP-HPLC was then used to evaluate the deprotection results (Figure 3-4). In all chromatograms a prominent peak at 26.2 min was visible, indicating the formation of one major product during solid-phase RNA synthesis and basic deprotection (Figure 3-4, 1a-1c). This signal is the base peak in the samples treated under protocols a) and b). In the sample from protocol c) a peak at 36.5 min represents the main product. Despite similar sample preparation and sample volume used for the HPLC analysis, the intensity of this peak is much higher. An explanation could be more efficient support-cleavage under protocol c) compared to the other protocols, yet incomplete cleavage of the oligo protecting group(s) (signal at 36.5 min). Concerning the homogeneity and purity of the oligo product, the deprotection protocol a) provided the best results.



**Figure 3-4.** Analytical HPLC profiles of the salicylaldehyde-containing RNA 15mer after the basic deprotection step (1a-1c) and after the additional TBS cleavage step (2a). HPLC conditions: 0–80% B in 45 min for 1, 0–40% B in 45 min for 2, detection at 260 nm.

For the subsequent removal of the TBS groups, the crude strands from a)-c) were treated with trimethylamine trihydrofluoride at 65 °C for 2.5 h. After quenching of the reagent and precipitation, the strands were analyzed by analytical RP-HPLC and MALDI-TOF. In the chromatograms of all three samples, the base peak was identical with a retention time of 22.6 min (Figure 3-4, 2a). Furthermore, the HPLC-profiles for the samples a), b) and c) became practically identical after silyl ether removal, indicating the formation of one main product. However, when the crude oligos were analyzed by MALDI-TOF mass spectrometry, a  $m/z = 4725$  was observed for all three samples (data not shown), which is not in accordance with the calculated mass of 4773.7 ( $M-H^-$ ). The lower mass of about 50 Da difference could neither be explained by depurination nor interruption of the solid-phase synthesis or common deprotection side reactions. Since the MALDI-TOF result for a synthesized control RNA strand without the salicylaldehyde was in agreement with the expected mass, the observed deviation was attributed to the artificial ribophosphoramidite. Also, the same problem was observed when other RNA strands containing a single salicylaldehyde base were synthesized by solid-phase synthesis.

Therefore, it was concluded that the free aldehyde of the ribophosphoramidite **88** was the reason for the observed difficulties during solid-phase synthesis and deprotection. In addition, this could also account for the encountered problems during 2'-OH protection. So, reprotection of the aldehyde before DMTr and 2'-OH protection was considered to be necessary, if the synthesis was repeated. At that time, further efforts towards the synthesis of an salicylaldehyde ribophosphoramidite **79** were postponed.

### 3.4 Transcription Experiments with STP and AldTP

Although RNA polymerases are closely related to DNA polymerases in function and structure, they differ significantly in their mode of action and regulation. In order to allow transcription, a double stranded promotor region is needed, which defines the starting point. The double strand is unwound and the RNA oligomerization starts by addition of a single nucleoside triphosphate. The so-called initiation of transcription is marked by low processivity and a special RNA polymerase conformation, the "open form". This conformation is essential for the promotor recognition, the double strand unwinding and transcription start. After addition of about 8-12 nucleotides the conformation changes to the "closed form", which is accompanied by an increase in processivity and an inability to bind the promotor again. Due to this conformation change and the affinity of the RNA polymerase to the promotor, "abortive initiation" occurs. This means transcription is started and is aborted after addition of a few nucleotides, because the polymerase rebinds the promotor. Only after a certain RNA length (usually 8-12 bases) is reached and the RNA polymerase changes to the "closed form", stable elongation can occur.<sup>[180]</sup>

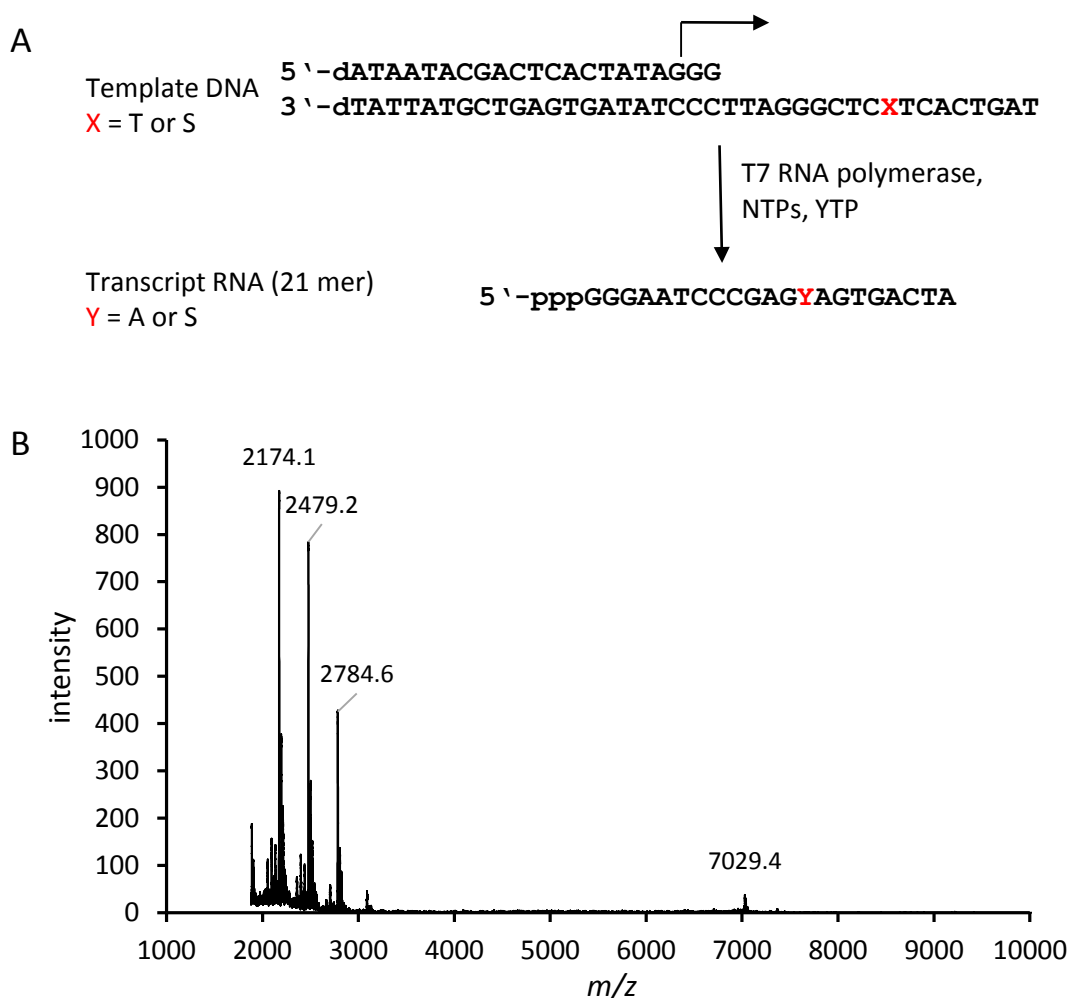
With the benzaldehyde **78** and salicylaldehyde ribonucleotide **1** in hand, DNA template strands for transcription experiments were designed. As the RNA polymerase from bacteriophage T7 was to be used in the transcription experiments, the DNA template contained a T7 early-stage promotor, which had been used for *in vitro* transcriptions with other artificial base pairs before.<sup>[181]</sup> In order to improve the transcription of the salicylaldehyde base, the position for the artificial base was positioned at +13, which is at the end of the abortive initiation region. Moreover, only a partial double stranded template was designed, in which the double stranded region is limited to the promotor region. This is advantageous, as only the antisense strand containing a single aldehyde base had to be synthesized and unwinding of the template during elongation is unnecessary. Thereby, any transcription stop caused by the mechanical resistance from the covalently linked salicylaldehydes was avoided.

Aldehyde and salicylaldehyde phosphoramidites were employed for the solid-phase synthesis of the template strand on a 1  $\mu$ mol scale. According to the trityl values, the performance of the aldehyde phosphoramidites was comparable to commercial phosphoramidites of the canonical bases (data not shown). Before cleavage from the solid-phase, an additional acidic deprotection step was needed to obtain the free aldehydes. The strands were carefully purified by RP-HPLC and their identity was verified by MALDI-TOF mass spectrometry. Acetal hydrolysis for the benzaldehyde base proved to be more difficult than for the salicylaldehyde. Using the exact same conditions for both modified strands, complete

deprotection was observed only for the salicylaldehyde-containing strand. For the benzaldehyde base, prolonged incubation with dichloroacetic acid was necessary to achieve similar deprotection results. The protocol was accordingly adapted.

Hybridization of each of these antisense strands with the short sense strand was achieved by heating of equivalent molar amounts to 95 °C followed by cooling at a rate of 2 °C/min to rt. For the transcription assays, 2 μM template, 400 μM nucleotides and varying amounts of amine additives and metal ions were mixed in the adequate buffer. The T7 RNA polymerase was added to start the reaction (Figure 3-5, A). After several hours incubation at 37 °C, the reactions were cooled on ice, dialyzed and analyzed by MALDI-TOF mass spectrometry.

For the transcriptional incorporation of the STP and AldTP opposite the salicylaldehyde in the antisense strand various conditions were screened. The canonical triphosphates were always used at 400 μM, the concentration of the aldehyde nucleotides was varied from 0.4-1.2 mM (0.4, 0.8, 1.2 mM). The amine concentration was varied from 0-140 mM (0, 10, 20, 30, 40, 80, 140 mM). In addition to the ethylenediamine, which was successfully used as the crosslinking amine in DNA replication,<sup>[182]</sup> methylamine and 1,2-diaminobenzene were studied as well. Upon imine formation between the template salicylaldehyde and the incoming (salicyl)aldehyde nucleotide, diaminobenzene would yield an extended conjugated system. This was considered favorable, because it would improve the important  $\pi$ -stacking interaction between the adjacent bases in the nascent RNA-DNA duplex. In contrast to DNA duplexes, which majorly adopt the B-conformation, RNA-DNA hybrids usually adopt an A-form like conformation.<sup>[183]</sup> As the A-conformation duplex has a wider major groove, sterical problems were expected upon crosslinking of the strands during imine formation. Methylamine would form two separate imines, thus creating the ligand scaffold without covalent linkage between the strands. Moreover, different metal ions were analyzed in combination with the three amines to allow transcription *via* metal base pair formation. Besides the Cu<sup>2+</sup> ion, Ag<sup>+</sup>, Mn<sup>2+</sup>, Fe<sup>2+</sup> and Fe<sup>3+</sup> were applied in transcription experiments involving STP or AldTP from 0.0-4.0 mM (0.0, 0.5, 1.0, 2.0, 3.0, 4.0 mM). Control transcriptions from a native template were conducted using these non-optimal conditions to allow differentiation between effects due to the conditions and the template.



**Figure 3-5.** (A) Principle of the *in vitro* transcription experiments. (B) MALDI-TOF spectrum of a control transcription using T7 RNA polymerase and 400  $\mu$ M NTPs. The main peaks at  $m/z = 2174$ -2785 correspond to abortive initiation products (6-8 mer), the full-length product is the 21-mer at  $m/z = 7029$ .

For analysis of the MALDI-TOF spectra from the transcription experiments, masses of the theoretically possible transcripts were calculated and then compared to the resulting peaks (Table 3-1). The MALDI-TOF spectrum of a transcription experiment from a native template under optimal conditions is shown in Figure 3-5, B. Despite the standard conditions, the most intense peaks originate from RNA strands that are 6-8 bases long and originate from abortive initiation. In comparison to these peaks, the signal at  $m/z$  7029, which corresponds to the full-length transcription product, appears negligible. However, the signal intensity is not directly proportional to the real RNA amounts in the sample. The MALDI-TOF mass spectrometry is more sensitive for shorter oligonucleotides, as these are more readily ionized.<sup>[184]</sup> Apart from this drawback, the possibility to directly assign peaks to a certain oligonucleotide and to recognize mutation of the artificial nucleoside in the transcript was invaluable.

**Table 3-1.** Calculated masses of possible transcripts.

length	Sequence	M	[M-H] <sup>-</sup>
21mer-A	pppGGGAAUCCCGAGAAGUGACUA	7034.93	7033.92
21mer-S	pppGGGAAUCCCGAGSAGUGACUA	7021.90	7020.89
20mer-S	pppGGGAAUCCCGAGSAGUGACU	6692.85	6691.84
19mer-S	pppGGGAAUCCCGAGSAGUGAC	6386.83	6385.82
18mer-S	pppGGGAAUCCCGAGSAGUGA	6081.74	6080.73
17mer-S	pppGGGAAUCCCGAGSAGUG	5752.69	5751.69
16mer-S	pppGGGAAUCCCGAGSAGU	5407.65	5406.64
15mer-S	pppGGGAAUCCCGAGSAG	5101.62	5100.61
14mer-S	pppGGGAAUCCCGAGSA	4756.57	4755.57
13mer-A	pppGGGAAUCCCGAGA	4440.55	4440.55
13mer-S	pppGGGAAUCCCGAGS	4427.52	4426.51
12mer	pppGGGAAUCCCGAG	4111.49	4110.48
11mer	pppGGGAAUCCCGA	3766.44	3765.43
10mer	pppGGGAAUCCCG	3437.39	3436.43
9mer	pppGGGAAUCCC	3092.34	3091.33
8mer	pppGGGAAUCC	2787.30	2786.29
7mer	pppGGGAAUC	2482.26	2481.25
6mer	pppGGGAAU	2177.22	2176.21
5mer	pppGGGAA	1871.19	1870.18

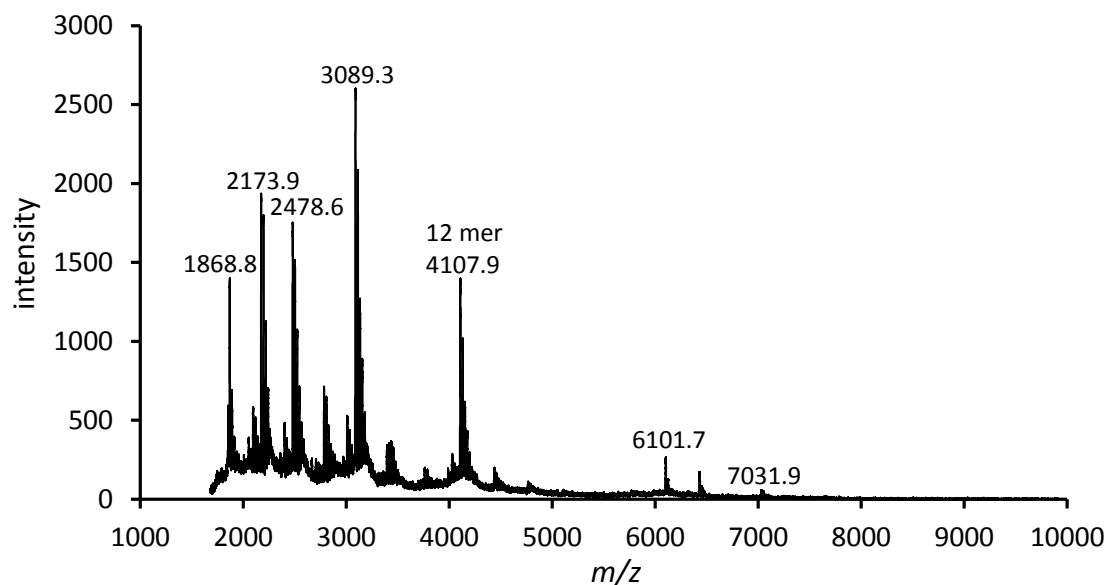
In order to decrease the number of experiments for the salicylaldehyde transcription condition screening, “optimal” concentration ranges for the additives were initially determined for a single additive individually. At the beginning, the triphosphate concentration was varied for STP only. The slight increase of abortive initiation products for 1.2 mM STP in the control transcription compared to 0.4 mM (data not shown) led to the decision to use the latter concentration for all subsequent experiments. Maybe a crosslinking reaction between the STP and an amine residue of the polymerase interferes with efficient transcription propagation. A similar strategy was applied for the amine additive. At first, only ethylenediamine was studied at different concentrations. Interestingly, the addition of up to 20 mM of ethylenediamine seemed to improve the full-length transcription product amount (data not shown). This is in accordance with observations made for other amines, which increase T7 RNA polymerase activity up to 12-fold.<sup>[185]</sup> Even 140 mM ethylenediamine did not deteriorate the transcription considerably. Therefore, amine additive concentrations of 0, 20 and 140 mM were chosen for all three amine additives in subsequent experiments. Likewise, Cu<sup>2+</sup> was selected as a starting point to narrow down the experiment number. As expected, an elevated Cu<sup>2+</sup> concentration of 4.0 mM almost stopped the production of the full-length transcript completely in the control experiments (data not shown). Even with 2 mM of Cu<sup>2+</sup> the common transcript pattern including full-length product was observed for the control reaction. Due to the T7 RNA polymerase preparation protocol, the final enzyme was provided in a 1 mM EDTA containing buffer, which complexed some of the added metal cation. The resulting concentration of available metal cation in the experiment was therefore lowered by 100-300  $\mu$ M based on the assumption of one Cu<sup>2+</sup> per EDTA. For the subsequent screen of alternative metal ions, the 2 mM concentration was hence selected as a starting point. Instead of performing almost 4000 transcription experiments to screen all conditions, these initial experiments decreased the amount to a feasible 100. Condition refinement, including reaction temperature,

incubation duration and concentration of the additives was to be performed for “interesting” results from these transcriptions.

No difference between  $\text{Cu}^{2+}$  and  $\text{Mn}^{2+}$  addition was observed in the experiments. Compared to these metal cations, equal amounts of silver and iron salts had a detrimental effect on transcription. More abortive initiation fragments were observed, and at 3 mM  $\text{Ag}^+$  no transcription products were observed at all (data not shown). Similar negative effects were found for diaminobenzene. This is in contrast to methylamine and ethylenediamine, which did not alter the transcription considerably.

The influence of the incubation time and temperature were studied as well. Short incubation durations (0.5 and 1.0 h) only provided MALDI spectra with low intensity, indicating low amounts of transcripts. Long incubations of 15 h or longer also led to a decreased signal intensity, which suggested residual RNase activity within the samples or metal-ion-mediated strand scission. An incubation time of 4-5 h was selected for almost all of the subsequent experiments, as it provided the best signal/noise ratio in MALDI and the best signals for full-length transcripts. For the reaction temperature, 37 °C incubation was compared to 20 °C and 40 °C. Almost no difference was observed between 20 °C incubation for 5 h compared to 37 °C. The same observation was made for the elevated temperature (40 °C, data not shown). Since 37 °C represents the optimal temperature for the T7 RNA polymerase activity, all subsequent experiments were performed at 37 °C, for 4-5 h. Initial experiments involving the salicylaldehyde-containing template, STP and  $\text{Cu}^{2+}$  and ethylenediamine showed the appearance of small amounts of 13 mer in addition to a strong 12 mer signal in the MALDI spectrum (data not shown). No full-length transcripts were found. Due to a poor signal/noise ratio, unambiguous determination whether STP incorporation had occurred or not was impossible. However, since similar MALDI spectra were obtained in experiments with AldTP and with only dNTPs, mutation of the salicylaldehyde position was considered. The addition of ethylenediamine and  $\text{Cu}^{2+}$  was necessary to force any transcription beyond the artificial nucleoside.





**Figure 3-6.** MALDI-TOF spectrum of a salicylaldehyde-containing transcription using T7 RNA polymerase, 400  $\mu$ M STP and NTPs, 1.4 mM  $\text{Cu}^{2+}$  and 20 mM ethylenediamine.

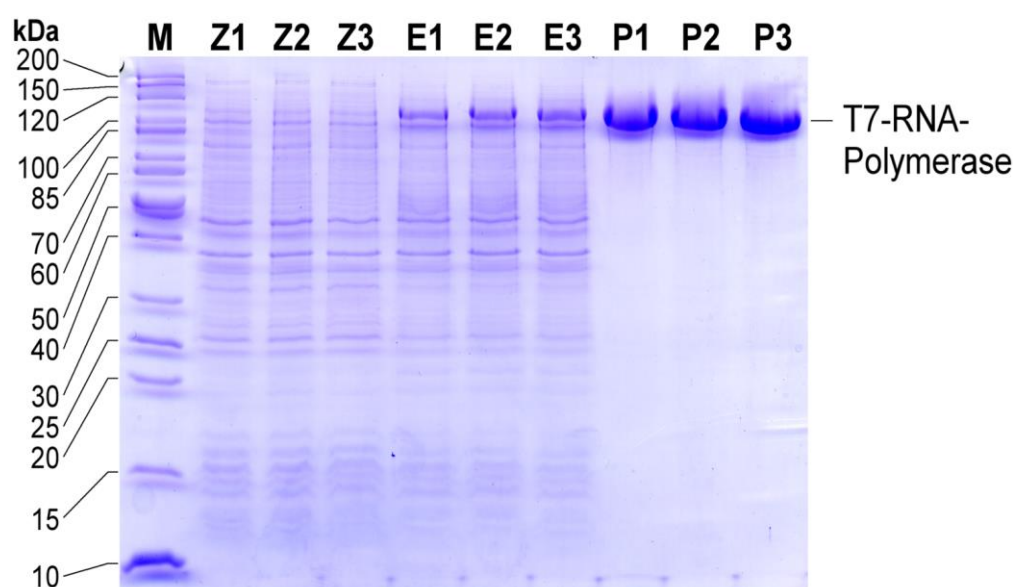
Finally a condition was found which allowed full-length transcription from a salicylaldehyde-containing template. By adding 1.4 mM  $\text{Cu}^{2+}$  and 20 mM ethylenediamine and incubation at 37 °C for 5 h, a tiny signal at 7031.9 was observed (Figure 3-6). When compared to the expected mass of 7021 (Table 3-1), it was obvious that the salicylaldehyde had been mutated. Calculations for alternative transcripts replacing the position by natural nucleosides revealed that adenosine was probably incorporated opposite the salicylaldehyde of the template (calc. 7033.9). A similar mutation had been observed before for the DNA salen pair, when no ethylenediamine and  $\text{Cu}^{2+}$  were present.<sup>[13]</sup> Again, the main signals in the spectrum were derived from abortive initiation fragments 1868-3089 (corresponding to 5-9 mers) and the 12 mer at 4107.9, which is directly before the artificial nucleoside in the template. This indicated the difficulties of the polymerase to transcribe past this position, despite the “misincorporation” of the undesired adenosine in position 13.

### 3.5 Transcription Experiments using T7 RNA Polymerase Mutants

In a last attempt to achieve transcription of the benzaldehyde or salicylaldehyde base, a biological approach involving mutation of the T7 RNA polymerase was envisioned. Unfortunately the use of other RNA polymerases was not feasible, as very few RNAPs are commercially available and each polymerase needs a different promotor and therefore a different template.

The T7 RNAP is well characterized and the crystal structure and data of several mutants are available in the literature.<sup>[186-187]</sup> The plasmid pBH161\_T7RNAP carrying the wild-type T7 RNA polymerase gene including a (His)<sub>6</sub>-tag was generously provided by Dr. Sabine Schneider. For mutagenesis, active site residues involved in the selection of nucleotides during transcription, were selected as a target. The mutation of the phenylalanine in position 644 and the histidine 784 were described to considerably increase

the rate of misincorporation of nucleotides and subsequent mispair extension.<sup>[187]</sup> In a PCR, mutated primers were used for site-directed mutagenesis of the T7 RNAP gene at the corresponding positions to alanine. The non-mutated methylated template was digested by *DpnI* and the reaction mixture including the mutated plasmid was subsequently transformed into *E. coli* DH5 $\alpha$ . Plasmids of colonies growing in the presence of a selection marker were sequenced to prove mutation at the desired position. T7 RNAP\_F644A and H784A were generated successfully. In contrast, all efforts to generate a double mutant combining both mutations, failed despite various primer designs and mutation strategies. Maybe without external induction by IPTG, the T7 RNAP is expressed at a background level and changes the transcriptome of the host cell. The increased amount of incorrect transcripts by the double mutant RNAP could lead to misfolded and nonfunctional proteins and therefore kill the host organism.



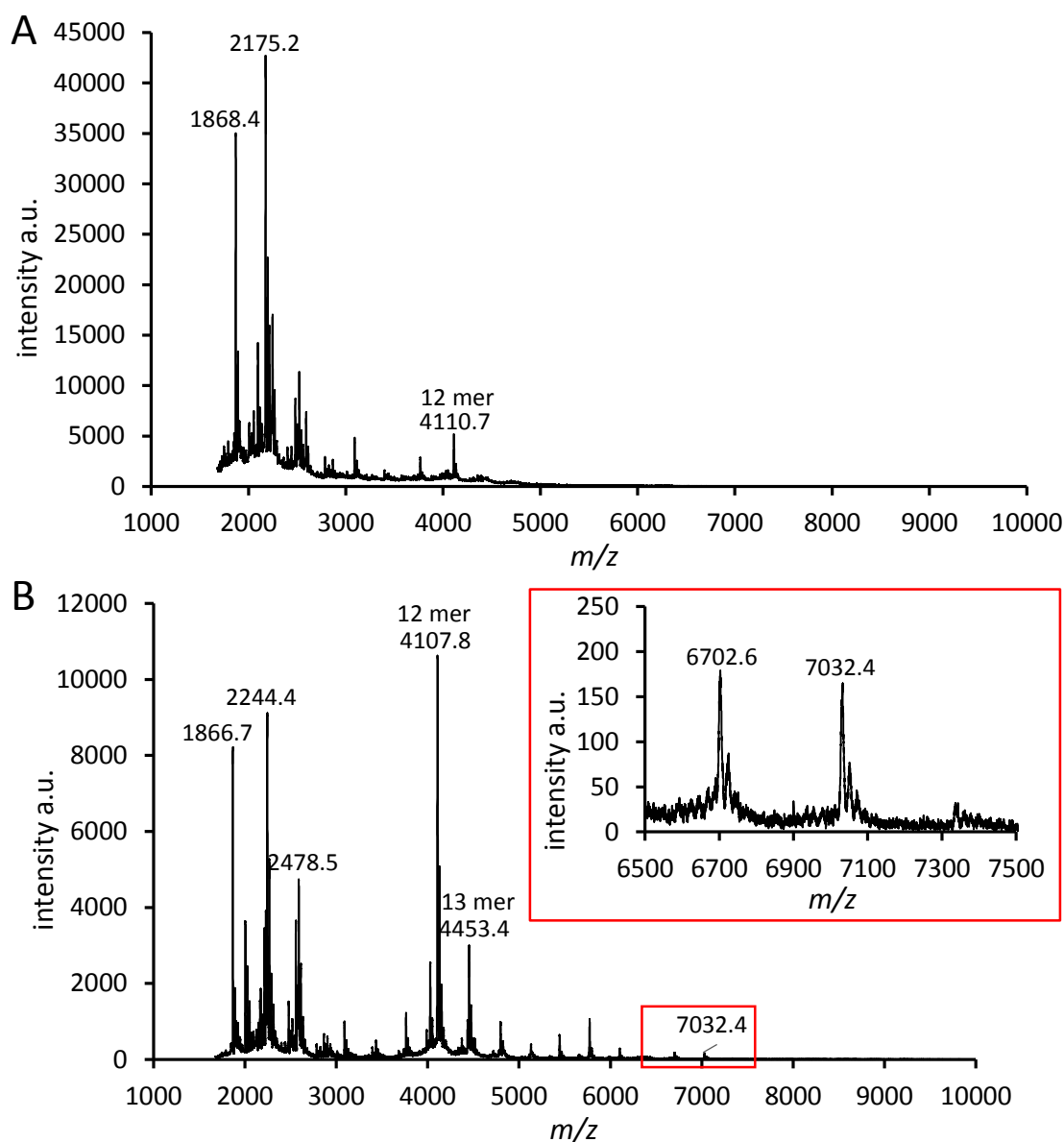
**Figure 3-7.** Coomassie stained SDS-PAGE of T7 RNA polymerase expression and purification. M = marker, Z1-3 = cell lysate of *E. coli* BL21 carrying plasmid pBH161\_T7RNAP before expression induction, E1-3 = cell lysate like in Z after induction using IPTG, P1-3 = purified T7 RNA polymerases. (1: wt, 2: F644A, 3: H784A).

The expression of the wild-type (1), the mutant F644A (2) and H784A (3) T7 RNAP were carried out in *E. coli* BL21 (Figure 3-7, E1-3). The T7 RNAP was purified from the soluble protein of the cell lysate by affinity purification with Ni-NTA beads. On average, 20-50 mg of T7 RNA polymerase were purified from 1 L culture in good purity as judged by SDS PAGE (Figure 3-7, P1-3).

In order to prove the activity of the purified enzymes, a test transcription was performed using the natural NTP set. Afterwards, transcription assays involving STP and AldTP and varied amounts of amines and metal ions were applied and the products were analyzed by MALDI-TOF mass spectrometry.

Differences in transcription experiments between the wild-type and the mutants were obvious. Surprisingly, the mutant F644A did not provide any full-length transcript, even when the “optimized” conditions were applied, which had given full-length product for the wild-type T7 RNA polymerase. Instead, transcription stopped before the incorporation opposite the artificial nucleoside (Figure 3-8, A).

For the mutant H784A also a peak from the 12 mer fragment was observed. This time the intensity was higher than for the abortive initiation fragments. In addition, a prominent peak of the 13 mer (4453) and a tiny peak of full-length transcript were found (7032). Both mass-to-charge ratios strongly suggested that under the applied conditions the salicylaldehyde in the template was transcribed into an A (cf. Table 3-1). In summary, no condition, (salicyl)aldehyde triphosphate or RNA polymerase mutant was found which allowed faithful transcription of the salicylaldehyde.



**Figure 3-8.** MALDI-TOF mass spectra from transcription experiments using T7 RNAP\_F644A (A) and T7 RNAP\_H784A (B) in the presence of 400  $\mu$ M STP and NTPs, 1.4 mM  $\text{Cu}^{2+}$ , 20 mM ethylenediamine and 2  $\mu$ M salicylaldehyde-containing template.

After all efforts to transcribe the salicylaldehyde nucleoside had failed, we decided to change the design of the base pair considerably. As metal ions reduce the compatibility with enzymes enormously, the new

concept was exclusively based on reversible imine chemistry. An amine base and an aldehyde base were to provide interstrand crosslinks, without the need for additives, like amines or metal ions. At the beginning, these new nucleosides had to be synthesized as phosphoramidites for their incorporation into DNA strands.

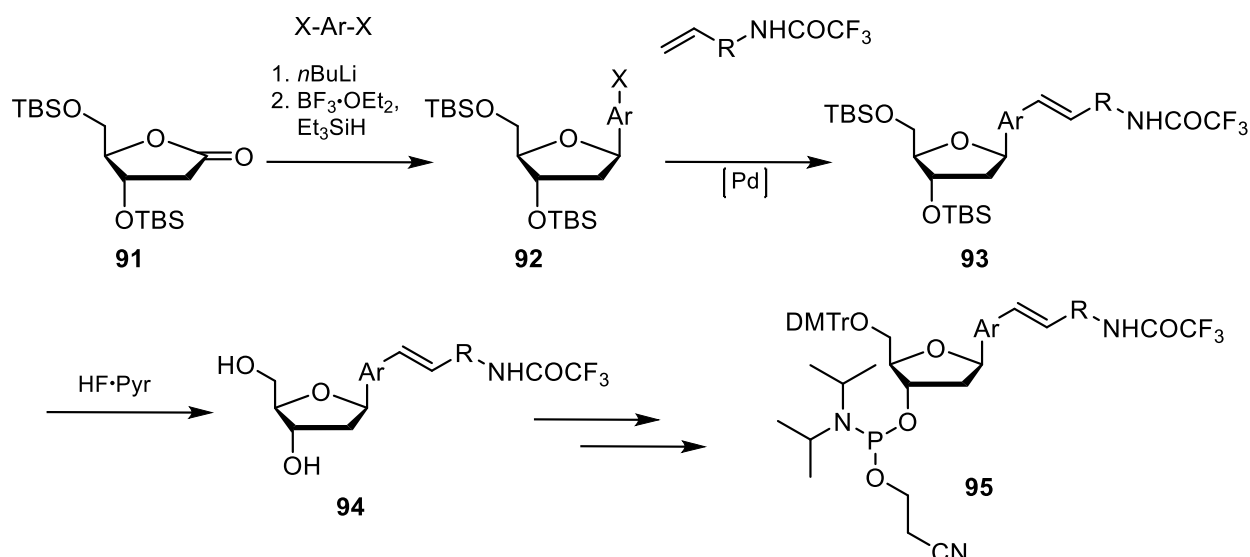
## Part II – A Covalent Base Pair

Although the salen base pair was selectively replicated in DNA, the need for the external additives ethylenediamine and Cu(II) ions is a major disadvantage for a broad application. The metal ions can denature enzymes and are toxic for living organisms at elevated concentrations. Therefore, we envisioned the concept of a base pair which solely interacts *via* a reversible covalent bond. To avoid the need for an external amine, a heterobase system comprising an amine- and an aldehyde-containing base was designed. As a starting point for the detailed design, the salicylaldehyde base was used as the aldehyde side in the beginning. The amine base design was based on conjugated systems which were isosteric to a salicylaldehyde with an ethylenediamine attached *via* an imine.

The design and synthesis of the amine base derivatives was performed by Dr. *M. Tomás-Gamasa*. During the project, a modular synthesis for amine base nucleosides and their corresponding phosphoramidites was established. This allowed the synthesis of a small library of amine base-containing strands that were analyzed against aldehyde base-containing strands in melting temperature experiments. Amine-aldehyde pairs which lead to an increase of the melting temperature were studied in more detail.

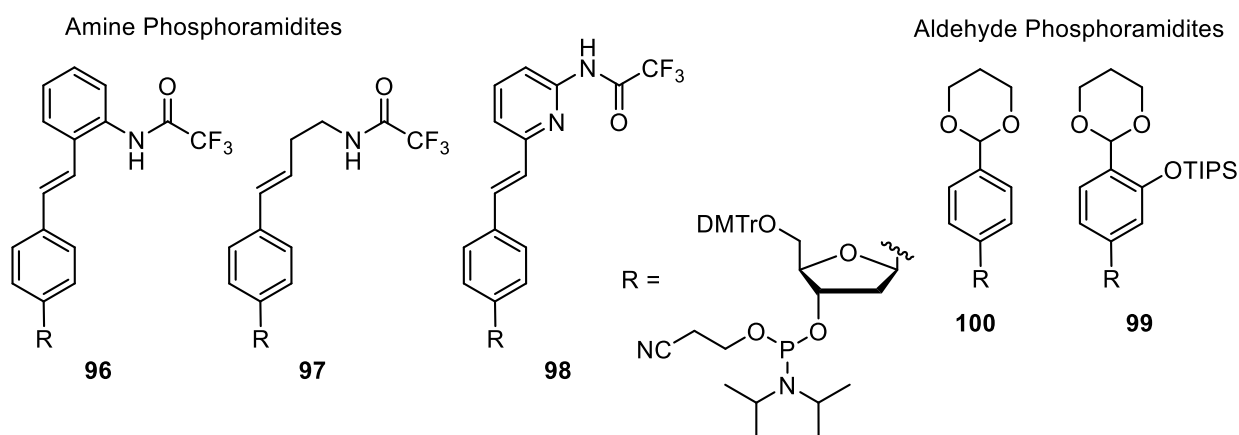
### 3.6 Overview of the Synthesized Amine and Aldehyde Phosphoramidites

For the synthesis of the amine base nucleosides a modular procedure was developed which involved the organometallic addition of a bishalide aromatic compound to a TBS-protected deoxyribolactone **91** (Scheme 3-12) and subsequent reduction of the intermediate hemiketal. This was basically identical to the method that had been used in the synthesis of the salicylaldehyde ribonucleotide, yet the protocol was adapted according to a procedure described in literature.<sup>[188]</sup> Various vinyl-containing amines were then cross-coupled to the halogen aromatic nucleoside **92** to access a small library of protected amine base nucleosides **93**. After deprotection of the sugar silyl ethers, the resulting nucleosides **94** were subsequently transformed into the corresponding phosphoramidites **95** using standard procedures. At the beginning, only 1,4-dibromobenzene was chosen as the aromatic halide and the vinyl-containing amines for the cross-coupling were varied. Three amine nucleoside phosphoramidites were prepared using this synthesis route: An aromatic amine **96**, an aliphatic amine **97** and a pyridine amine **98** phosphoramidite (Figure 3-9). Upon imine formation with the (salicyl)aldehyde nucleoside, the aromatic amine and the aliphatic amine were supposed to adapt a structure with a similar sterical demand compared to the salen complex. Since the amino group in the pyridine amine is shifted by one position compared to the other two amine nucleosides, the geometrical necessities for imine formation could be studied.



**Scheme 3-12.** Modular synthesis of the amine phosphoramidites for solid-phase synthesis.

The synthesis of the salicylaldehyde base nucleoside and its phosphoramidite was established in the group.<sup>[12]</sup> Thus, freshly prepared dS phosphoramidite **99** was directly used for the synthesis of strands for melting temperature experiments. In addition, the synthesis of a benzaldehyde base nucleoside phosphoramidite was designed based on the established routes. Compared to the salicylaldehyde it lacks the *ortho*-hydroxyl group and thereby was used to determine its importance for the stability of the imine. The synthesis was described and performed by Meng Su.<sup>[41]</sup> Briefly, 4-bromobenzaldehyde was acetal protected, lithiated and added to TBS-protected deoxyribose lactone **91**. The resulting hemiketal was reduced to the protected nucleoside using Lewis acid conditions. Finally the TBS groups were removed and the nucleoside was transformed to the DMTr-protected phosphoramidite **100**. Altogether, three amine and two aldehyde phosphoramidites were prepared for solid-phase synthesis of strands for melting temperature analyses.



**Figure 3-9.** Overview of the synthesized amine and aldehyde phosphoramidites. DMTr = 4,4'-dimethoxytrityl, TIPS = triisopropylsilyl.

### 3.7 Melting Temperature Analysis of Strands Containing Amine and Aldehyde Bases

With the three amine base phosphoramidites and the two aldehyde base phosphoramidites in hand, DNA strands for melting temperature analysis were synthesized. The synthesis of the salicylaldehyde containing strand was performed as described by *C. Kaul* and *G. Clever*.<sup>[12-13]</sup> Based on this protocol, which only needed an additional acidic cleavage step during deprotection, the strand containing the aldehyde nucleoside was prepared. According to the trityl values, no difference in performance was observed compared to the commercial phosphoramidites. After the deprotection procedure, the mass of the desired benzaldehyde-containing strand and of the acetal-protected form was found in MALDI-TOF samples. Obviously, the acetal of the benzaldehyde nucleoside is more stable than the acetal of the salicylaldehyde nucleoside. Therefore, the acidic cleavage step after the solid-phase synthesis was prolonged. The aldehyde strands were purified by RP-HPLC, their identity was verified by MALDI-TOF mass spectrometry and the total concentration was determined by UV-Vis.

The synthesis of the amine base-containing strands needed more adaption of the protocols. Initial efforts to apply the standard solid-phase DNA synthesis procedure suffered from a deprotection problem. Although all amine base phosphoramidites were synthesized with trifluoroacetyl protecting groups on the amine residue, the strands which were isolated after the deprotection were found to possess a single acetyl group. As the standard phosphoramidites are completely deprotected under the applied conditions, we concluded that the trifluoroacetyl group had been replaced by the acetyl group during a capping step. Deprotection of the TFA group during solid-phase synthesis could be promoted by the acidic “deblock solution” and thereafter the amine was acetyl-protected in the capping step. Unfortunately, no suitable deprotection condition was found which allowed removal of the acetyl group without strand decomposition, except for the pyridine amine. Therefore, the capping step after incorporation of the amine base was omitted in the solid-phase DNA synthesis. This enabled the successful synthesis of the amine-containing strands for the melting temperature experiments. Although the capping steps after about half of the solid-phase synthesis were omitted, no considerable decrease in yield and coupling efficiency was observed. Strands containing the aromatic amine, the aliphatic amine and the pyridine amine nucleoside were synthesized. After deprotection, purification by RP-HPLC and identity confirmation by MALDI-TOF mass spectrometry, the concentration was determined by UV-Vis. For a comprehensive overview of the strands that were synthesized and utilized in various combinations for the melting temperature studies see Table 5-5, p. 149.

**Table 3-2.** Overview of the synthesized amine and aldehyde base-containing strands for melting temperature experiments and the purchased control strands. n.d. = not determined

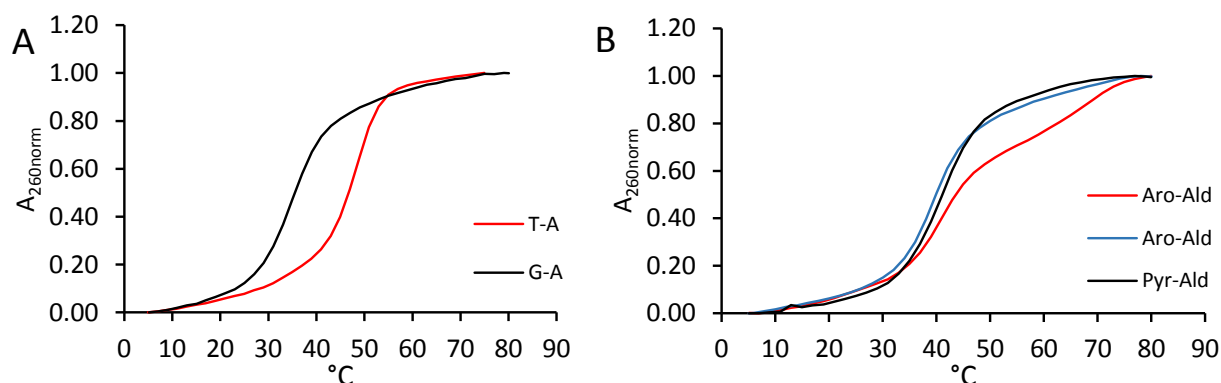
Entry	Name	Oligonucleotide Sequence 5'-3'	Mass (calc.)	Mass (found)
1	Benzaldehyde 1	TAC AAC AAldT AAT GTG	4555.0	4552.7
2	Benzaldehyde 2	CAC ATT AAldT GTT GTA	4537.0	4535.2
3	Salicylaldehyde 1	TAC AAC AST AAT GTG	4571.0	4569.6
4	Salicylaldehyde 2	CAC ATT AST GTT GTA	4553.0	4551.8
5	Aromatic Amine	CAC ATT AAroT GTT GTA	4625.9	4624.0
6	Aliphatic Amine	CAC ATT AAlit GTT GTA	4578.1	4576.0
7	Pyridine Amine	CAC ATT APyrT GTT GTA	4627.1	4625.5
8	Control 1	CAC ATT AAT GTT GTA	4565.0	n.d.
9	Control 2	TAC AAC ATT AAT GTG	4574.1	n.d.
10	Control 3	TAC AAC AGT AAT GTG	4599.1	n.d.

Equal amounts of complementary strands were hybridized and the melting temperatures were determined by following the 260 nm UV-signal upon several cycles of heating and cooling. At the beginning, the benzaldehyde strands (Table 3-2, Entry 1+2) were studied in the presence and absence of external amines and metals. For comparison, control strands possessing only natural bases were measured containing a central match or mismatch pair. The combination of the benzaldehyde strands had a melting temperature ( $T_M$ ) of 37 °C, which is only slightly higher than the 35 °C of a duplex having a central C-A or A-G mismatch. This  $T_M$  was neither changed by the addition of methylamine, 1,2-diaminobenzene, ethylenediamine nor by additional metal ions like  $\text{Cu}^{2+}$ ,  $\text{Mn}^{2+}$  and  $\text{Ag}^+$ . It was therefore concluded that a stable imine formation for the benzaldehyde nucleoside is not possible under the applied conditions. Next, combinations of a benzaldehyde strand and a salicylaldehyde strand (Table 3-2, Entry 1+4 and 2+3) were analyzed with and without amines and metal ions. Under all conditions the  $T_M$  was determined as 40 °C. This is in between the melting temperature of a mismatch (35 °C) and a match (48 °C). Once again, the results indicated that the benzaldehyde nucleoside cannot form a stable imine bond with the provided amines in contrast to the salicylaldehyde self-pair.

Then, amine base-containing strands were combined with (salicyl)aldehyde-containing strands. The combination of the aliphatic amine-containing strand with the benzaldehyde strand had a melting temperature of only 37 °C (Table 3-3, entry 5). A slight increase to 39 °C was observed for the pyridine amine and even 42 °C for the aromatic amine opposite the benzaldehyde (Figure 3-10, B, black and red). Moreover, in the case of the aromatic amine and the benzaldehyde, the curves for the annealing and melting were not superimposable. Instead, a  $T_M$  of 38 °C was determined for the annealing (Figure 3-10, B, blue). All combinations of the amine strands with the benzaldehyde strand did result in a melting temperature decrease compared to the  $T_M$  of 48 °C for the control duplex with the central T-A pair (Figure 3-10, A, red). Instead, the melting temperatures are slightly above the 35 °C, which were determined for

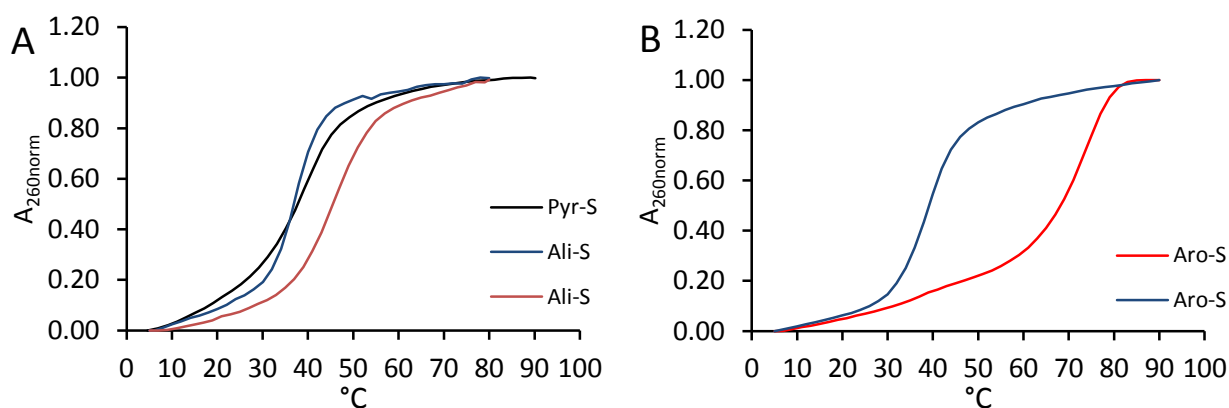


the G-A or C-A mismatch. This indicates that the combination of the aliphatic amine or the pyridine amine with the benzaldehyde generates a situation similar to a mismatch, with the imine bond not established. For the aromatic amine with a melting temperature in between a match and a mismatch, it was assumed that the imine is partly formed. Upon close observation of the corresponding melting curve (Figure 3-10, B, red), actually two melting events can be detected, one at 42 °C and the other at 68 °C. We therefore suggest that only for the DNA fraction which has the higher melting temperature the imine is formed. The other fraction behaves like a mismatch.



**Figure 3-10.** Melting profiles of control duplexes (A) and combinations of the benzaldehyde with an aromatic amine and pyridine amine containing strand (B). Note that the blue (annealing) and the red (melting) curve in B were obtained from the same double strand.

Next, the three amine base containing strands were combined each with the salicylaldehyde strand. The combination of the strand containing the pyridine amine with the salicylaldehyde had a melting temperature of 38 °C (Figure 3-11, A, black). For the aliphatic amine opposite the salicylaldehyde two melting temperatures, at 37 °C and 44 °C for the annealing (blue) and the melting (red) were determined. As the curves from cooling and heating were not superimposable, different pathways for annealing and heating were implied. The delayed reaction of the system and the dependence of the reaction outcome from the past state (here temperature) is called hysteresis.<sup>[189]</sup> In comparison to the control, both combinations had a decreased melting temperature. However, for the aliphatic amine the interaction to the salicylaldehyde, resulted in only a minor decrease by 4 K, which is 9 K higher than the  $T_M$  of 35 °C for the mismatch. This indicated successful imine formation for the combination. In accordance to the observations made for the benzaldehyde base, the melting temperature increase for the aromatic amine opposite the salicylaldehyde was the highest. A  $T_M$  of 79 °C was determined for the heating and 38 °C during annealing. Thus, the hysteresis effect was more distinct compared to the two other amine bases. The strong increase of the melting temperature indicated that a stable interstrand imine bond was formed.



**Figure 3-11.** Melting profiles of double strands with a central (A) salicylaldehyde pyridine amine (black), salicylaldehyde aliphatic amine (red, blue) and (B) salicylaldehyde aromatic amine combination (red, blue). The blue curves are from annealing, the red curves are from melting if hysteresis was observed.

In order to study the observed hysteresis behaviour in the case of the salicylaldehyde and the aromatic amine, melting curves were measured at various pH values and at faster and slower cooling/heating rates. Initially, the cooling/heating rate was decreased from 1.0 K/min to 0.1 K/min. The  $T_M$  for the heating decreased from 79 to 62 °C, whilst the  $T_M$  determined for the cooling increased from 38 to 46 °C. Thus, the hysteresis decreased from 41 to 16 K, which indicates that a slow process is the reason for the observed effects. A further increased rate of 5 K/min had only a minor consequence for the  $T_M$  (79 and 39 °C) compared to the starting condition.

**Table 3-3.** Determined melting temperatures of the DNA duplexes described in this chapter. The names are chosen according to the central base pair, the sequence of the strand combinations can be found in Table 3-2. All values below were determined at pH 9 and a rate of 1 K/min cooling/heating.

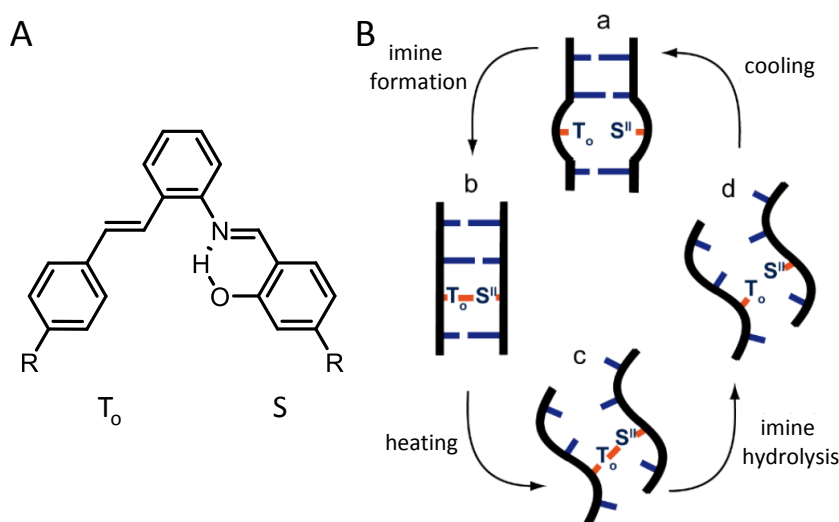
Entry	Name	Strand Combination	$T_M$ [°C]
1	T-A match	Control 1 and Control 2	48
2	G-A mismatch	Control 1 and Control 3	35
3	Ald-Ald	Benzaldehyde 1 and Benzaldehyde 2	37
4	Sal-Ald	Salicylaldehyde 1 and Aldehyde 2	40
5	Ald-Ali	Benzaldehyde 1 and Aliphatic Amine	37
6	Ald-Pyr	Benzaldehyde 1 and Pyridine Amine	39
7	Ald-Aro	Benzaldehyde 1 and Aromatic Amine	38, <sup>[a]</sup> 42 <sup>[b]</sup>
8	Sal-Ali	Salicylaldehyde 1 and Aliphatic Amine	37, <sup>[a]</sup> 44 <sup>[b]</sup>
9	Sal-Pyr	Salicylaldehyde 1 and Pyridine Amine	38
10	Sal-Aro	Salicylaldehyde 1 and Aromatic Amine	38, <sup>[a]</sup> 79 <sup>[b]</sup>

[a]  $T_M$  for the annealing, [b]  $T_M$  for the melting

Then, the pH was varied at a constant cooling/heating rate of 1.0 K/min in order to study the influence of protonation states on the hysteresis. Based on the conditions that had been applied for the salicylaldehyde metal base pair, the starting pH of 9 was selected in a CHES buffered solution for all described

experiments. At this pH the melting temperature increase of the salicylaldehyde metal base pair had been found highest. Also for the S:T<sub>o</sub> (salicylaldehyde:aromatic amine) pair, the highest melting temperature of 79 °C was determined under these conditions. When the pH was decreased to 7, the T<sub>M</sub> for the melting decreased to 68 °C. At the same time the T<sub>M</sub> determined for the hybridization during cooling increased from 38 to 46 °C. Thus, the observed hysteresis decreased from 41 to 22 K, which indicates that protonated/deprotonated residues are involved in the melting/reannealing process of the S:T<sub>o</sub> pair.

The smallest hysteresis of only 6 K was observed at pH 7 and a rate of 0.1 K/min. At this condition melting temperatures of 49 and 55 °C were measured for the hybridization and the melting, respectively. Therefore, both T<sub>M</sub> are above 48 °C, which was observed for the control duplex containing a central A-T pair. This means that selective binding between the salicylaldehyde and the aromatic amine is possible, when the system has enough time to adjust a suitable conformation.



**Figure 3-12.** Proposed (A) H-bond stabilized imine form and (B) hysteresis model of the T<sub>o</sub>:S pair.

Based on the findings for the S:T<sub>o</sub> pair, a model was proposed which might explain the hysteresis effect (Figure 3-12, B). The central hypothesis is a slow rate of imine formation and hydrolysis. When the duplex is heated, the canonical base pairs melt, but the central S:T<sub>o</sub> crosslink stabilizes the duplex (c) until the imine bond is hydrolyzed (d). During cooling, the natural bases might already adapt Watson-Crick base pairing, yet the central unformed crosslink represents a typical mismatch (a), which slowly reacts to the imine (b).

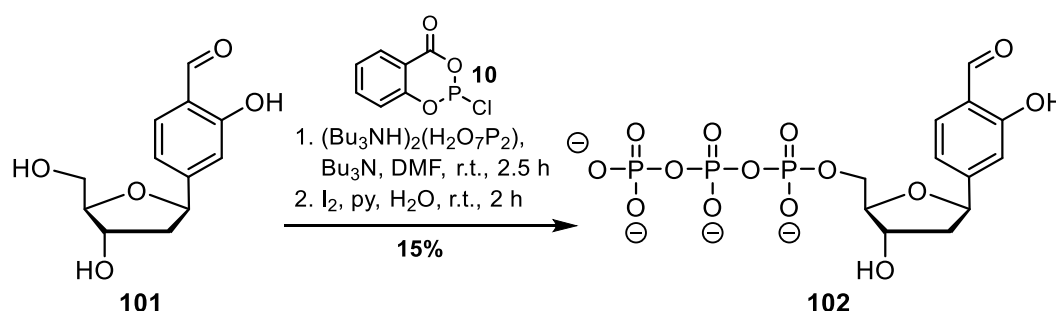
From all aldehyde-amine combinations in the melting temperature experiments, only the interaction between the aromatic amine T<sub>o</sub> and the salicylaldehyde S had a stabilizing effect on the duplex compared to the natural base pairs. The *ortho*-hydroxyl group adjacent to the aldehyde in the salicylaldehyde was determined as an important factor for the imine stability, as the combination of the aromatic amine and the benzaldehyde base destabilizes the duplex. An intramolecular H-bond from the hydroxyl group to the imine, which increases the stability of the reversible covalent bond (Figure 3-12, A), was therefore

---

suggested. Moreover, the importance of  $\pi$ -stacking and conformational space for base pair formation was highlighted by comparing the interaction of the aliphatic amine and the aromatic amine with the salicylaldehyde. Although both amine bases possess the amine in almost the same distance from the 1'-carbon, the increase of the melting temperature and thus a stable imine formation was only observed for the aromatic amine. We assume that the  $\pi$ -stacking to the directly adjacent bases in the duplex aligns the aromatic amine in a position which can react preferably with the aldehyde. Also the loss of entropy upon binding is smaller, as the fully conjugated aromatic amine is more rigid compared to the aliphatic amine. The observed hysteresis, especially for the aromatic amine with the salicylaldehyde, was an indication for the slow imine formation and hydrolysis. Maybe this characteristic can be exploited in future work to construct a nanodevice with a time-dependent memory or DNA origami with a time-dependent shape-memory.

### 3.8 Synthesis of a Salicylaldehyde and an Aromatic Amine Deoxyribonucleotide

After observation of the stabilizing interaction between salicylaldehyde **S** and the aromatic amine **T<sub>o</sub>** in the melting temperature experiments, enzymatic incorporation of the artificial nucleotides by DNA polymerases was studied. This could allow to generate long covalently linked DNA strands for especially stable nanostructures that are not accessible by conventional solid-phase DNA synthesis. In order to analyze the enzymatic acceptance, primer extension experiments were planned. The corresponding DNA templates were synthesized with the existing phosphoramidites, which had been used for the melting strand synthesis before. Only the deoxyribonucleotides of the artificial bases were missing at that time.

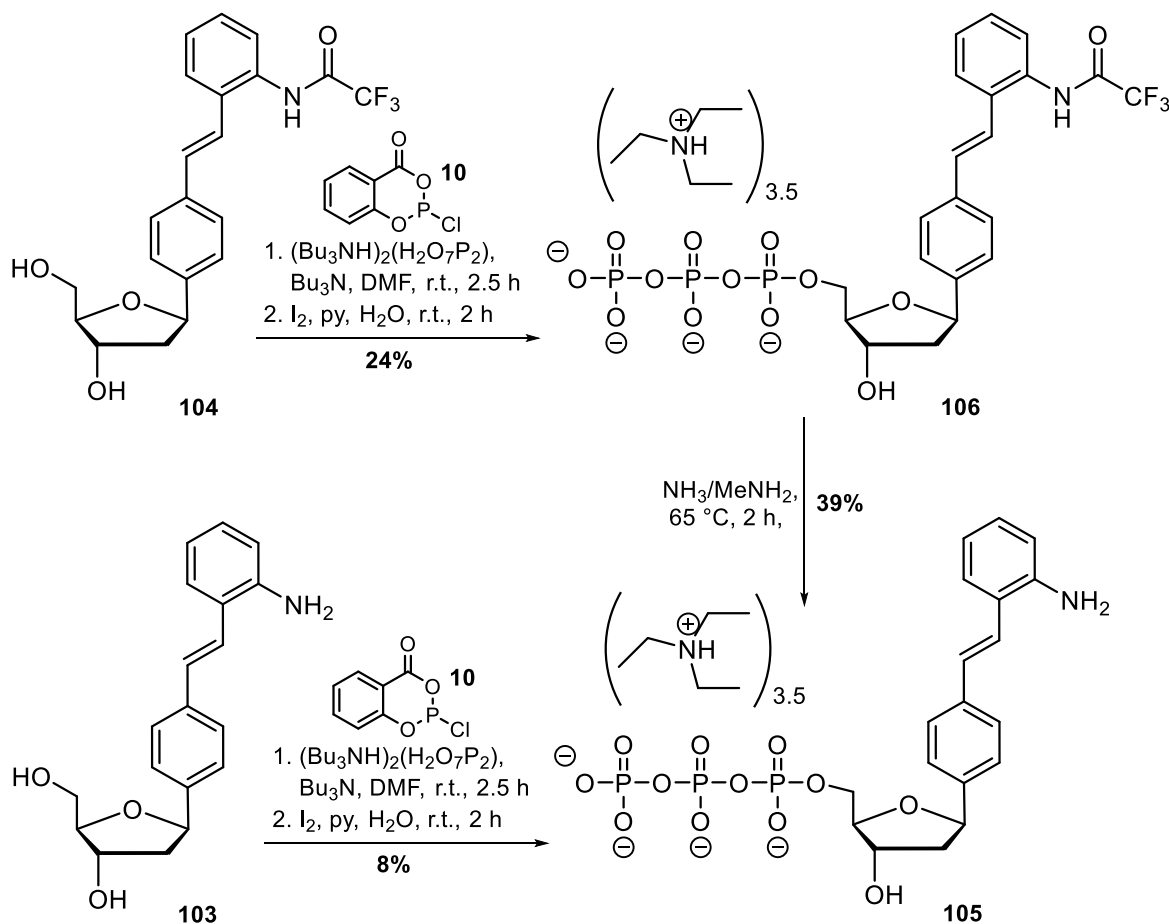


**Scheme 3-13.** Synthesis of the salicylaldehyde deoxynucleoside 5'-triphosphate **102** (dSTP).

The synthesis of the salicylaldehyde deoxynucleoside **101** and its phosphoramidite **99** was designed and established by *G. Clever*. Also dSTP had been synthesized by *C. Kaul* and *M. Wagner* before, following the famous *Ludwig-Ruth* triphosphate reaction protocol (1.3 Chemical (Tri)phosphorylation Methods, p. 4).<sup>[48-49]</sup> However, the RP-HPLC purification was tedious and the dSTP was obtained in only 2% yield.<sup>[13]</sup> Protected salicylaldehyde nucleoside was still present from the performed phosphoramidite synthesis and provided a one-step access to the salicylaldehyde deoxynucleoside. Having made good experiences with the more recent triphosphorylation method in several syntheses, we decided to apply the reaction to the salicyl deoxynucleoside **101** as well.<sup>[64]</sup> Satisfyingly, after twofold RP-HPLC purification, the dSTP **102** was obtained in 15% yield (Scheme 3-13). The NaCl-ethanol precipitation step during the reaction workup of the triphosphate immensely improved the purity of the crude triphosphate product and therefore simplified the purification. HRESI mass spectrometry in combination with  $^{31}\text{P}$  and  $^1\text{H}$  NMR proved the formation of the 5'-dSTP in accordance with published data.<sup>[13]</sup> In analytical RP-HPLC a very broad peak was observed for the dSTP (data not shown), which hampered a meaningful estimation of the purity. This behavior is based on the special property of the salicylaldehyde deoxynucleoside and has been observed before.<sup>[182]</sup> As no additional signals to the signals of 5'-dSTP were observed in the  $^1\text{H}$  and  $^{31}\text{P}$  NMR spectra, the purity was estimated to be at least 95%.

The synthesis of the aromatic amine nucleoside **103** (**T<sub>o</sub>**) and its phosphoramidite was designed and established by *Dr. M. Tomás-Gamasa*. Generously, some TFA-protected aromatic amine nucleoside **104** and the completely deprotected nucleoside **103** were provided for the synthesis of the aromatic amine 5'-triphosphate. To ensure a successful synthesis two triphosphate reactions were setup in parallel. In the first

case, the TFA-protected aromatic amine **104** was used to avoid chemoselectivity problems during the triphosphorylation reaction and was to be saponified to give the aromatic amine deoxynucleoside triphosphate **105** (dT<sub>0</sub>TP) in a subsequent reaction. In the second case, the aromatic amine **103** was directly applied to the triphosphorylation reaction without any protecting group (Scheme 3-14).

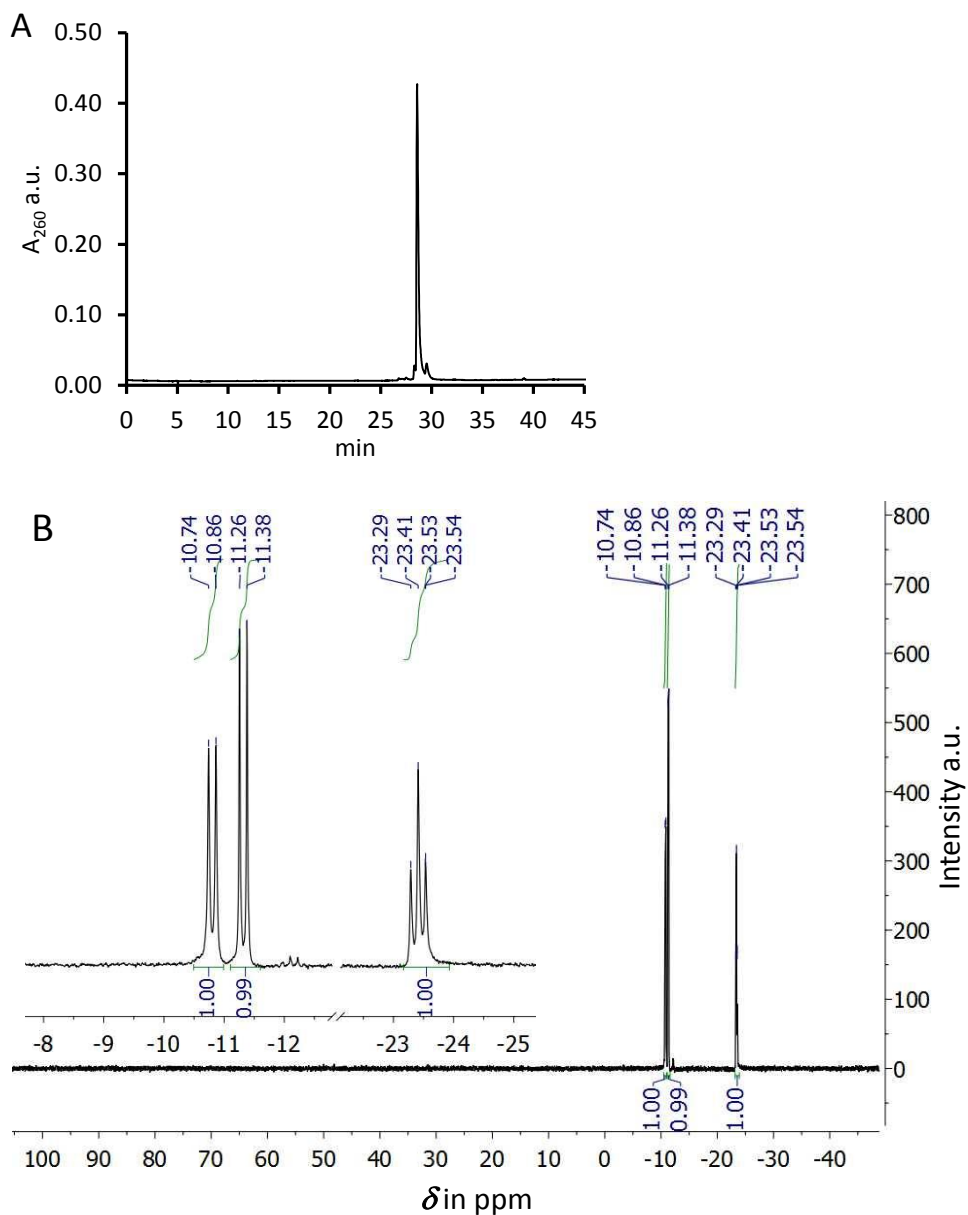


**Scheme 3-14.** Synthesis of the aromatic amine deoxynucleoside 5'-triphosphate dT<sub>0</sub>TP **105** by two different routes.

Surprisingly, initial experiments using the established triphosphate protocol did not provide the desired product. The freshly purchased 2-chloro-1,3,2-benzodioxaphosphorin-4-one (**10**) was completely oxidized as judged by <sup>31</sup>P NMR. Therefore, the phosphite reagent **10** was synthesized from salicylic acid and phosphorus trichloride according to a described procedure, using only small amounts of toluene as a solvent.<sup>[190]</sup> As PCl<sub>3</sub> is very toxic, corrosive and also prone to oxidation, distillation of the purchased chemical was avoided. Also utmost care was taken to prevent any leakage of PCl<sub>3</sub> and also developing HCl gas during reflux, by passing the cooled gases through an aqueous saturated solution of NaHCO<sub>3</sub>. All steps of the reaction and the workup were performed under an atmosphere of argon to prevent the oxidation of the P(III) species. In the <sup>31</sup>P NMR spectra a prominent peak at 148.2 ppm proved the presence of the desired P(III) species. Another peak was observed at -3.5 ppm at about 20% of the total <sup>31</sup>P species, which originates from the unwanted P(V) form of the reagent. This demonstrates the sensitivity of the reagent for oxidation but could also be due to POCl<sub>3</sub> present in the

employed  $\text{PCl}_3$  due to longterm storage. As the crude phosphite reagent was acceptably pure, distillation of the almost colorless solid was omitted in subsequent reactions.

Gratifyingly, the so-synthesized salicyl phosphite **10** yielded the desired triphosphates in both setups when employed in the reaction protocol (Scheme 3-14). After two-fold RP-HPLC purification the amine protected aromatic triphosphate **106** was obtained in 24% yield. The alternative route from the aromatic amine nucleoside gave dT<sub>o</sub>TP **105** in 8% isolated yield and also 4% 3'-triphosphate side product. Despite the lower yield in the case of the unprotected amine, no phosphoramidate product, which could originate from the reaction of the aromatic amine with the preformed triphosphate reagent, was isolated. Instead, much nucleoside was reisolated according to analytical RP-HPLC. It was assumed that although the phosphoramidate side product is formed in the reaction, it is mostly hydrolyzed upon workup and purification. For the deprotection of the amine protected triphosphate **106** to the dT<sub>o</sub>TP **105**, conditions similar to protocols from DNA solid-phase synthesis strand cleavage were used. This yielded 39% of the desired dT<sub>o</sub>TP **105** after RP-HPLC purification. The low yield is probably due to hydrolysis of the triphosphate at the elevated pH during the deprotection procedure. NMR spectra unambiguously proved the identity of the 5'-dT<sub>o</sub>TP **105**, and analytical RP-HPLC verified the high purity of the product (Figure 3-13).



**Figure 3-13.** (A) Analytical RP-HPLC trace of the purified dT<sub>o</sub>TP **105** at 260 nm detection using a gradient from 0–70% B in 45 min. (B)  $^{31}\text{P}$  NMR spectrum of the purified dT<sub>o</sub>TP **105**.

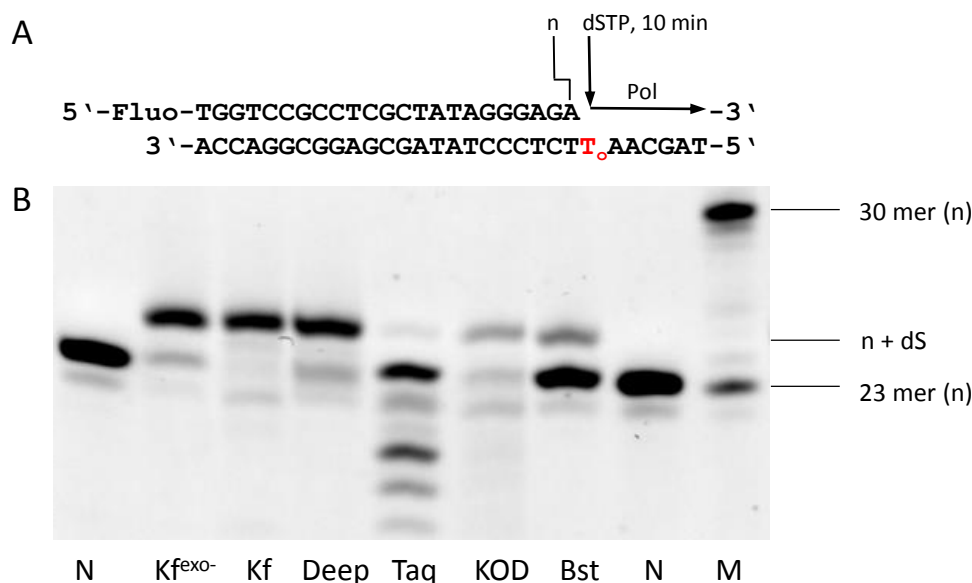


### 3.9 Primer Extension Experiments with dSTP and dT<sub>o</sub>TP

Several compounds were necessary for the primer extension experiments involving the artificial salicylaldehyde base dS and aromatic amine dT<sub>o</sub>. Apart from the triphosphates, dSTP **102** and dT<sub>o</sub>TP **105**, template strands containing each artificial base in the n+1 position were prepared from the corresponding phosphoramidites. The strands were synthesized and deprotected according to the protocols established for the strands of the melting temperature studies (3.7). After RP-HPLC purification and confirmation of the identity by MALDI-TOF, the purity was assessed by analytical RP-HPLC. Oligonucleotides with a purity greater than 95% were used for the primer extension experiments. In addition, control templates and 5'-fluorescein labeled primers were purchased from METABION. Control templates were unmodified in the n+1 position, and thus allowed evaluation whether the applied experimental conditions were suited for the polymerases. The 5'-fluorescein labeled primer enabled analysis of the primer extension reaction *via* its fluorescence after denaturing polyacrylamide gel electrophoresis (PAGE). Sequences of synthesized and purchased strands for primer extensions presented within this work can be found in Table 5-6, p. 149. Utmost care was taken to avoid any source of external primary or secondary amine, like e.g. from the buffer used for the primer extension, that could interfere with the imine formation between dS and dT<sub>o</sub>. Therefore, tris(hydroxymethyl)amino-methane (Tris) and spermidine containing buffers were replaced by custom buffer mixtures. The buffering agent was replaced by *N*-cyclohexyl-2-aminoethanesulfonic acid (CHES for pH 8-9) and 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES for pH 7-8). The DNA stabilizing compound spermidine was omitted, as it is not crucial for DNA polymerase function. For further details on buffer composition and primer extension conditions see Table 5-1, p. 147 and 5.4.15, p. 157.

#### 3.9.1 Primer Extensions with dSTP

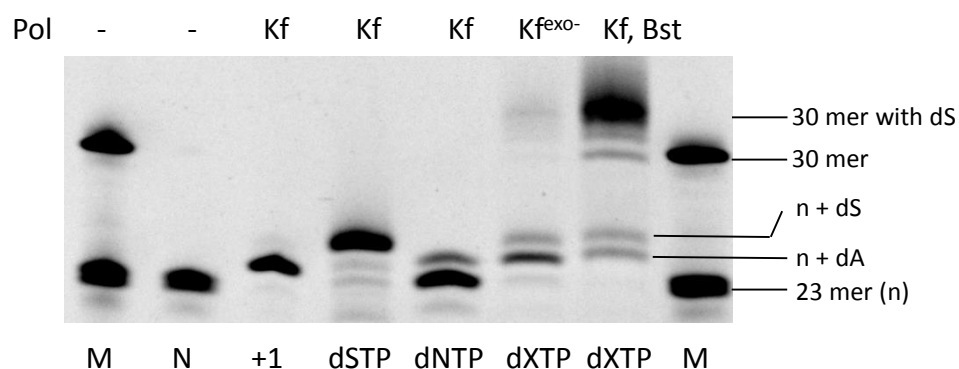
In the beginning, various polymerases were analyzed for their ability to incorporate the salicylaldehyde nucleotide (dSTP) opposite the templating aromatic amine (T<sub>o</sub>) base in single nucleotide incorporations (SNI). Therefore, a template hybridized from the primer P1a and the T<sub>o</sub>-containing strand P1b was incubated with different DNA polymerases in the presence of 400 μM dSTP (Figure 3-14, A). Several polymerases were found capable of forming the base pair under the applied conditions of the experiment to a different extent. However, only Klenow fragment (KF) managed quantitative conversion of the primer as judged after denaturing PAGE (Figure 3-14, B). The exonuclease deficient forms of KF and Deep vent<sup>®</sup> showed some remaining non-elongated primer. For *Bst* Pol I and KOD XL only traces of the +1 elongated primer were observed (Figure 3-14, B). Instead of elongation products, degradation of the primer was observed in the case of One *Taq*<sup>®</sup>. Notably, the exonuclease activity of KF did not consider the artificial base as a mismatch, as exclusive n+dS product formation was observed.



**Figure 3-14.** (A) Template and principle of the primer extension experiment. (B) Denaturing PAGE from SNI experiments in the presence of 400  $\mu\text{M}$  dSTP using different polymerases. M is the marker and N is the negative control. Polymerases:  $\text{KF}^{\text{exo-}}$  = Klenow fragment exonuclease deficient, KF = Klenow fragment, Deep = Deep vent, Taq = One Taq, KOD = KOD XL, Bst = *Bst* Pol I.

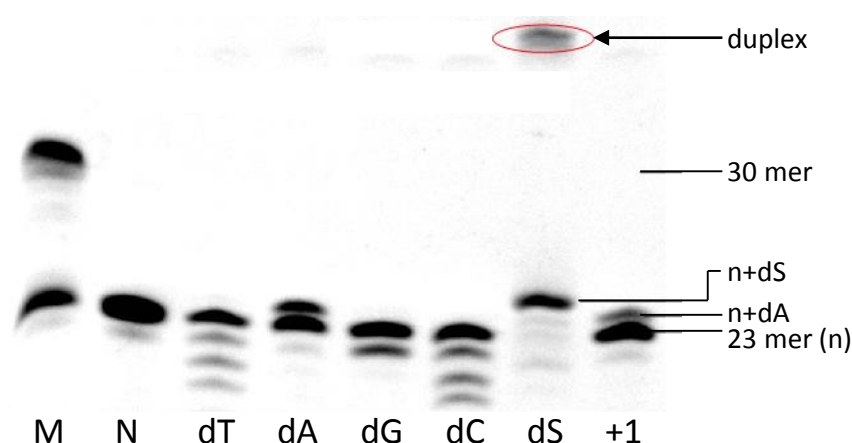
Encouraged by the good incorporation by KF, full elongation of the primer was attempted. When all five dXTPs were applied in primer extensions to the  $\text{T}_0$  containing template, only for  $\text{KF}^{\text{exo-}}$  a very faint band of fully elongated primer was found (Figure 3-15, dXTP). Most of the other polymerases including KF yielded the +1 elongation product or just the primer (data not shown). Extension beyond an incorporated artificial base pair is known to be a major roadblock in artificial base pair design.<sup>[191-192]</sup> Next, a two-step procedure was applied in which first the dSTP was incorporated by KF followed by addition of the natural dNTPs and a second polymerase. So, each polymerase was studied for its ability to elongate the dS containing primer, independently of its capability to selectively form the S: $\text{T}_0$  pair. In this way, KF and  $\text{KF}^{\text{exo-}}$  only provided traces of the completely elongated primer and mostly the dS containing primer (data not shown). However, *Bst* Pol I was able to further elongate after dSTP incorporation by KF. It is noteworthy that *Bst* Pol I was found unable to form S: $\text{T}_0$  efficiently (cf. Figure 3-14) and thus did not provide any full-length product, when 5 dXTPs were applied. By combining the polymerases KF and *Bst* Pol I, complete elongation was achieved in the presence of 5 dXTPs (Figure 3-15). Only faint additional bands were observed. These were assigned to the n+dS product, but also n+dA misincorporation opposite d $\text{T}_0$ , and even further elongation was detected. It is assumed that KF incorporates the dSTP at 37 °C within 10 min, then the temperature increase to 60 °C denatures this polymerase and *Bst* Pol I continuous elongation.

The special property of the salicylaldehyde base to change the migration speed in the PAGE could be used to distinguish between correct elongation and mutasynthesis by dA. When only dNTPs were added to a primer extension by KF, a faint band of the +1 product was observed (Figure 3-15, dNTP), which migrates faster than the primer elongated by dS, (Figure 3-15, dSTP).



**Figure 3-15.** Denaturing PAGE from primer extension experiments using 1 pmol template, 2U polymerase and varying nucleotide compositions. M is the marker, N is the negative control and +1 is the control elongated by a single nucleotide. Single nucleotide insertion by KF (10 min, 37 °C) for dSTP or dNTP. (Attempted) full elongation by KF<sup>exo-</sup> (6 h, 37 °C) and a mixture of KF and *Bst* Pol I (i. 10 min, 37 °C, ii. 6 h, 60 °C) in the presence of 5 dXTPs. Polymerases: KF<sup>exo-</sup> = Klenow fragment exonuclease deficient, KF = Klenow fragment, Bst = *Bst* Pol I.

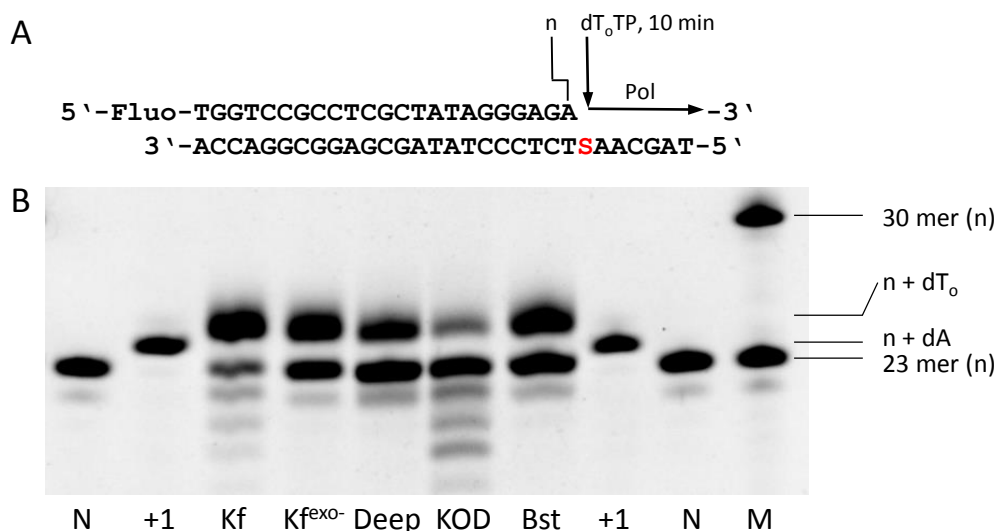
In order to exclude mutation of the artificial base position by a natural nucleotide, the selectivity of the dSTP incorporation by KF was studied in single nucleotide insertions. Therefore, each of the five nucleotides was provided in separate reactions with the T<sub>0</sub> template. While dSTP completely yielded the +1 elongation, only for dATP some +1 elongation was observed after denaturing PAGE. This indicated that mutation of dS to dA could be possible. For dCTP, dGTP and dTTP even some degradation of the primer was detected (Figure 3-16). Moreover, an additional slowly migrating band was observed for the dSTP incorporation, which indicated crosslink formation between the elongated primer and the template. Despite the denaturing conditions of the PAGE some double strand was observed. This demonstrated the stabilizing effect of the imine connection between the aldehyde and the amine base.



**Figure 3-16.** Denaturing PAGE from SNI experiments for the T<sub>0</sub> template in the presence of 200 μM dTTP, dATP, dGTP, dCTP and 400 μM dSTP using KF polymerase (10 min, 37 °C).

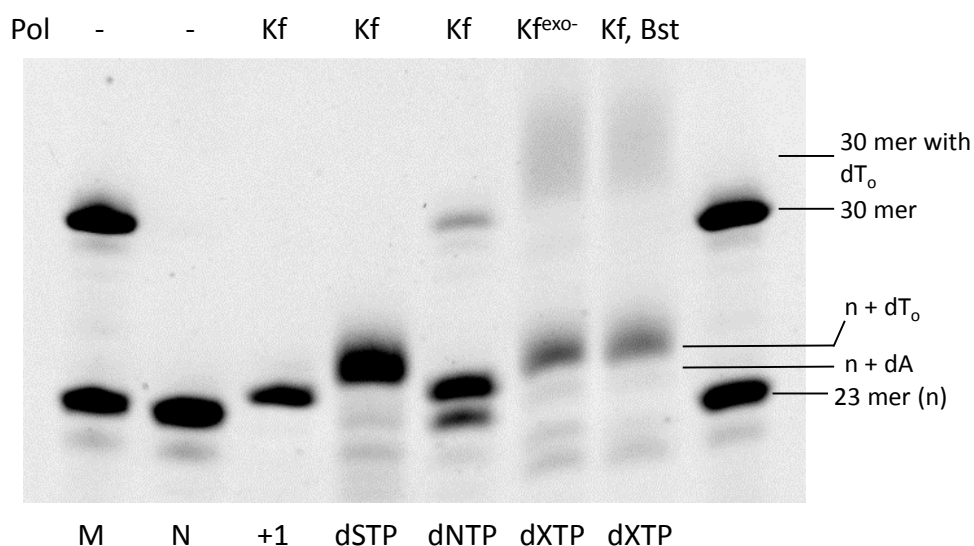
### 3.9.2 Primer Extensions with dT<sub>o</sub>TP

The experiments for the incorporation of dT<sub>o</sub>TP opposite a templating dS (oligo P1a and P1c, Table 5-6, p. 149) were performed analogously to the primer extensions in the converse case. The results from polymerase screening in SNI providing only 400  $\mu$ M of dT<sub>o</sub>TP were different from the previous experiments for the dSTP. None of the studied polymerases managed quantitative +1 elongation within 10 min, although some extension was observed for KF, KF<sup>exo-</sup>, Deep vent <sup>exo-</sup> and *Bst* Pol I (Figure 3-17). Again, KF proved to be most suited for the incorporation of the artificial nucleotide.



**Figure 3-17.** (A) Template and principle of the primer extension experiment. (B) Denaturing PAGE from SNI experiments in the presence of 400  $\mu$ M dT<sub>o</sub>TP using different polymerases. M is the marker, N is the negative control and +1 is the positive control. Polymerases: KF = Klenow fragment, KF<sup>exo-</sup> = KF exonuclease deficient, Deep = Deep vent, KOD = KOD XL, Bst = *Bst* Pol I.

Quantitative single nucleotide elongation for the dT<sub>o</sub>TP was achieved by increasing the incubation time to 20 min using KF. This condition was also applied to the first step in the two-step procedure aiming at full primer elongation. In the second step various polymerases were analyzed for their ability to incorporate dNTPs to the dT<sub>o</sub>-elongated primer. However, only for KF<sup>exo-</sup> a very faint band of full-length primer was observed (Figure 3-18). All other studied polymerases were found completely unable to elongate after dT<sub>o</sub>TP incorporation. Even the polymerase mixture that was able to provide full-length product in the case of dSTP opposite the templating dT<sub>o</sub>, did not provide more than traces of the desired product. The reasons for the observed differences in primer extensions between incorporation of dT<sub>o</sub>TP opposite dS and dSTP opposite dT<sub>o</sub> were enigmatic at that point. Also the encountered difficulties to provide full-length product at all, were proof for the shortcomings of the dT<sub>o</sub>:dS pair in enzymatic replication.



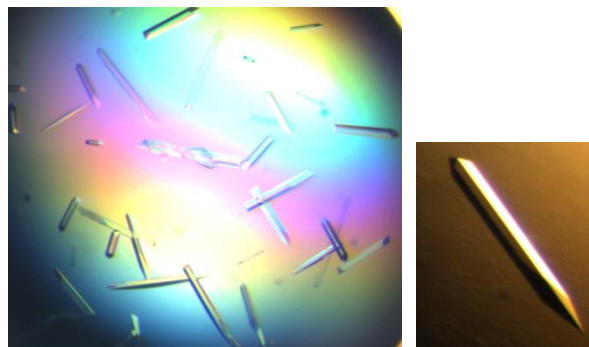
**Figure 3-18.** Denaturing PAGE from primer extension experiments using 1 pmol template, 2U polymerase and varying nucleotide compositions. M is the marker, N is the negative control and +1 is the control elongated by a single nucleotide. Single nucleotide insertion by KF (20 min, 37 °C) for dSTP or dNTP. Attempted full elongation by KF<sup>exo-</sup> (6 h, 37 °C) and a mixture of KF and *Bst* Pol I (i. 10 min, 37 °C, ii. 6 h, 60 °C) in the presence of 5 dXTPs. Polymerases: KF<sup>exo-</sup> = Klenow fragment exonuclease deficient, KF = Klenow fragment, Bst = *Bst* Pol I.

### 3.10 Co-crystallization of *Bst* Pol I with S:T<sub>0</sub> Containing DNA

In order to gain insight into the structural properties of the covalent base pair, DNA strands containing the S:T<sub>0</sub> were crystallized together with a DNA polymerase. For this reason the DNA polymerase from *Geobacillus stearothermophilus* (*Bst* Pol I) was chosen, because it allowed full elongation for dSTP opposite dT<sub>0</sub> and had been crystallized successfully in the group before. Expression and purification were performed according to the established protocols and the purified enzyme was generously provided by *Barbara Steigenberger*. Concomitantly, DNA strands were synthesized containing the artificial nucleobases in various positions to allow crystallization of preinsertion- (+2), insertion- (+1) and postinsertion (-5) complexes (Table 3-4) and were carefully purified by RP-HPLC. For the design of the template DNA 5'-overhangs of 3-4 bases were chosen, as this is the normal substrate for the polymerase in contrast to the alternative blunt end DNA. The obtained strands were hybridized to a final concentration of 500 µM and one volume was mixed with an equal volume of the purified *Bst* Pol I 10 mg/mL. This equates a molar DNA:protein ratio of about 3:1, which after 1 h incubation on ice was used for the crystallization setups. In the case of the preinsertion complex dSTP was added to a final 660 µM before the incubation step to allow formation of a ternary complex between polymerase, DNA and incoming nucleotide. Crystals were grown using the hanging drop vapor diffusion technique in 0.1 M MES (pH = 5.8), 46-49% of 4 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and 2-3% 2-methyl-2,4-pentanediol (MPD). Gratifyingly, within 3-14 days the polymerase-DNA crystals (Figure 3-19) grew at 18 °C for all three DNA templates. They were harvested after three weeks, transferred into a cryo-solution and frozen until the X-ray experiment. Beam-line measurements were kindly performed by *Dr. Sabine Schneider*. The crystal structure was solved by *Dr. Markus Müller* using molecular replacement and the existing structure with the pdb code: 2XY5. Details regarding data collection and structure refinement can be found in 5.4.14 p. 154.<sup>[41]</sup>

**Table 3-4.** Overview of the DNA strands and templates that were used for co-crystallization with *Bst* Pol I.

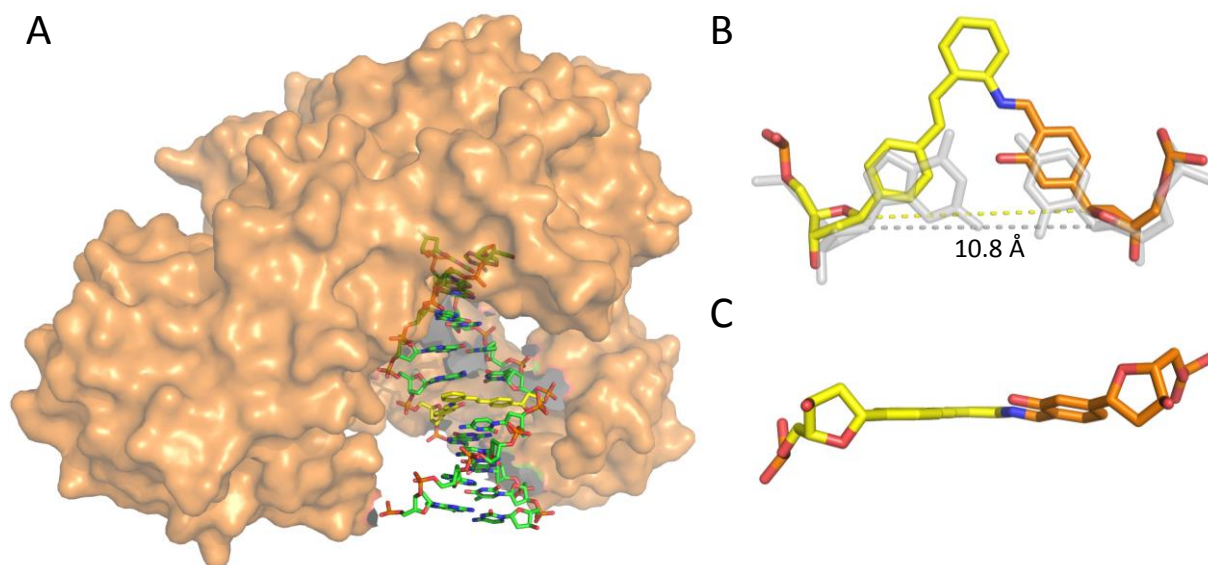
Entry	Name	DNA strands	Template
1	Preinsertion	5'-CACT <sub>0</sub> CGAGTCAGGCT-3' 5'-AGCCTGACTCG-3'	5'-CACT <sub>0</sub> CGAGTCAGGCT-3' 3'-GCTCAGTCCGA-5'
2	Insertion	5'-CACT <sub>0</sub> CGAGTCAGGCT-3' 5'-AGCCTGACTCG <b>S</b> -3'	5'-CACT <sub>0</sub> CGAGTCAGGCT-3' 3'- <b>S</b> GCTCAGTCCGA-5'
3	Postinsertion	5'-ATGCGACCT <sub>0</sub> TCCCT-3' 5'-AGGGA <b>S</b> GGTC-3'	5'-ATGCGACCT <sub>0</sub> TCCCT-3' 3'-CTGG <b>S</b> AGGGA-5'



**Figure 3-19.** Crystals of DNA-containing *Bst* Pol I depicted as grown in mother liquor after 12 days in polarized light. (Left) drop containing a DNA with the S:T<sub>o</sub> base pair in the postinsertion template (-5) at 0.1 M MES (pH = 5.8), 3% MPD and 47% of 4 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, (right) close-up view of a single crystal.

Well-diffracting crystals were obtained for all three setups which scattered with adequate resolution (about 2.0 Å) to perform molecular replacement. However, in the initial structure before the structure refinement it was noticed that the DNA template was orientated with the blunt end towards the active site instead of the expected 5'-overhang as intended. Unfortunately, this meant that for the preinsertion and insertion template setups the interesting DNA part was pointing away from the enzyme. In the region so distant from the polymerase active site, the DNA strand is more flexible and therefore cannot result in a well-resolved structure. Thereby, the data collected for the crystals from the insertion and preinsertion template were rendered useless. Reasons for the observed preference of the blunt end over the 5'-overhang in all crystal structures could result from the incubation step of the polymerase together with the DNA strand at 0 °C. Maybe an incubation step at rt or even elevated temperatures, like 60 °C, which are close to the enzyme activity optimum can resolve the issue. A definite solution to this problem is the design of the DNA template as a palindrome.

Luckily, the position of the S:T<sub>o</sub> base pair in the postinsertion complex was basically in the centre of the DNA duplex, even when the blunt end was in the active site of *Bst* Pol I. Therefore, the refined structure including the S:T<sub>o</sub> base pair inside the DNA duplex could be solved for this setup. The base pair is so far away from the active site that the polymerase simply serves as a crystallization scaffold for the DNA.



**Figure 3-20.**<sup>[41]</sup> (A) Crystal structure of *Bst* Pol I in complex with a T<sub>0</sub>:S crosslink containing DNA. (B) The T<sub>0</sub>:S crosslink overlaid with a canonical dG:dC base pair. The C1'-C1' distance of 10.8 Å for the unnatural crosslink is identical with the distance observed for the canonical base pair. (C) Rear view of the T<sub>0</sub>:S crosslink, showing a twist out of plane. T<sub>0</sub> is depicted in yellow, S in orange.

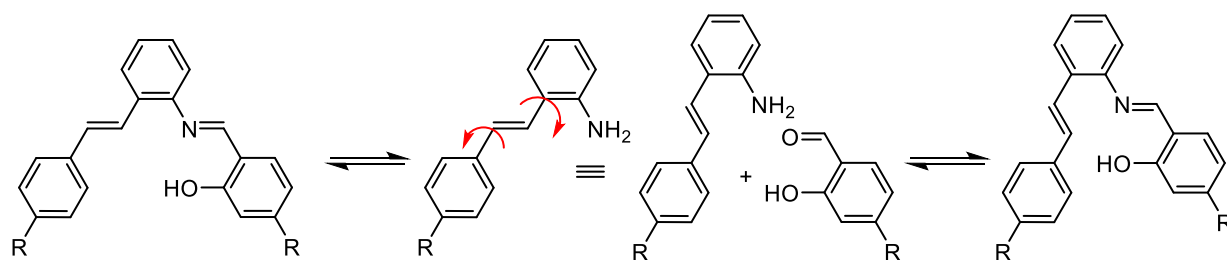
The structure shows that the T<sub>0</sub>:S bases face each other and that they are covalently linked *via* the expected imine interaction. An overlay with a structure of a canonical dG:dC base pair at this position proves that the T<sub>0</sub>:S crosslink provokes only small structural perturbations in the duplex (Figure 3-20, B). The distance of 10.8 Å between the C1' in the deoxyriboses of the artificial crosslink is identical to the natural G:C pair (10.8 Å). In contrast, the previously reported metal- and ethylenediamine mediated S homobase pair, has a distance of 11.4 Å.<sup>[13]</sup> The improved fit of the T<sub>0</sub>:S crosslink is achieved because T<sub>0</sub> adopts a conformation which was unexpected, with the stilbene double bond rotated away from the counterbase. It could be that the need to populate this rotameric conformation is the reason for the sluggish full elongation after dT<sub>0</sub>TP incorporation opposite a S template. The T<sub>0</sub>:S crosslink is not perfectly planar, with the S slightly twisted from the T<sub>0</sub> plane (Figure 3-20, C). The oxygen atom from the S base and the imine bond are placed in the same plane, like a bicyclic system trying to avoid the steric clash with the alkene.

Based on the X-ray results a model for the enzymatic incorporation of T<sub>0</sub> opposite the templating dS was proposed, which might explain the observations made in primer extension experiments (Scheme 3-15). When the dT<sub>0</sub>TP is incorporated opposite the templating dS the imine bond forms in a conformation which is expected for the S:T<sub>0</sub> nucleoside dimer in solution. This “open” conformation of the base pair disturbs proper DNA duplex formation, as it is too wide. The isomerization to the “closed” form, which we observed in the crystal structure, could occur *via* hydrolysis of the imine, rotation next to the alkene and reformation of the imine. As the imine hydrolysis is slow and the isomerization requires the dT<sub>0</sub> nucleobase to adapt an unfavorable conformation, primer extensions stop after dT<sub>0</sub>TP incorporation.

Interestingly, the incorporation of dSTP opposite a templating dT<sub>0</sub> is much easier and the primer can be fully elongated. Therefore, it is assumed that the canonical bases in the template strand directly adjacent to



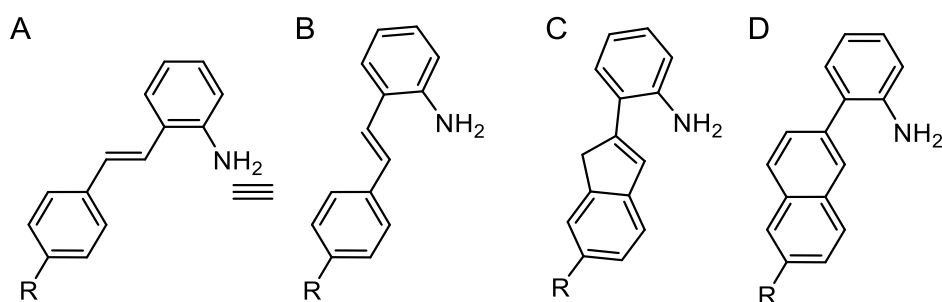
the dT<sub>o</sub> support the “closed” conformation through  $\pi$ -stacking. By stabilizing this conformation, e.g. by chemical modifications and redesign of the T<sub>o</sub> nucleoside, the acceptance by the polymerase and the incorporation efficiency should be improved.



**Scheme 3-15.** Proposed model for a possible conformational isomerization in a growing DNA strand. On the left is the conformation we expect in solution and upon single nucleotide insertion in unproductive primer extensions, on the right is the conformation observed in the crystal structure which is isosteric to a canonical base pair.

### 3.11 Chemical Efforts to Improve the Enzymatic Incorporation Efficiency

In the crystal structure we observed that the improved fit of the T<sub>o</sub>:S base pair was due to an unfavorable conformation of the aromatic amine base with the *trans*-stilbene rotated away from the center. As this is the less favorable conformation in solution, we hypothesized that this could also account for the poor processivity of the DNA polymerase in primer extension experiments. By “locking” the aromatic amine nucleoside through a cyclic system that is analogous to the less favorable conformation, an improved acceptance by polymerases and an increase in melting temperature was expected for this “improved” nucleoside compared to the initial T<sub>o</sub>:S pair. Possible molecular analogs of the unfavorable *trans*-stilbene conformation are an indene and a naphthalene derivative (Figure 3-21). As the indene derivative is difficult to synthesize, the naphthalene was selected as a replacement.

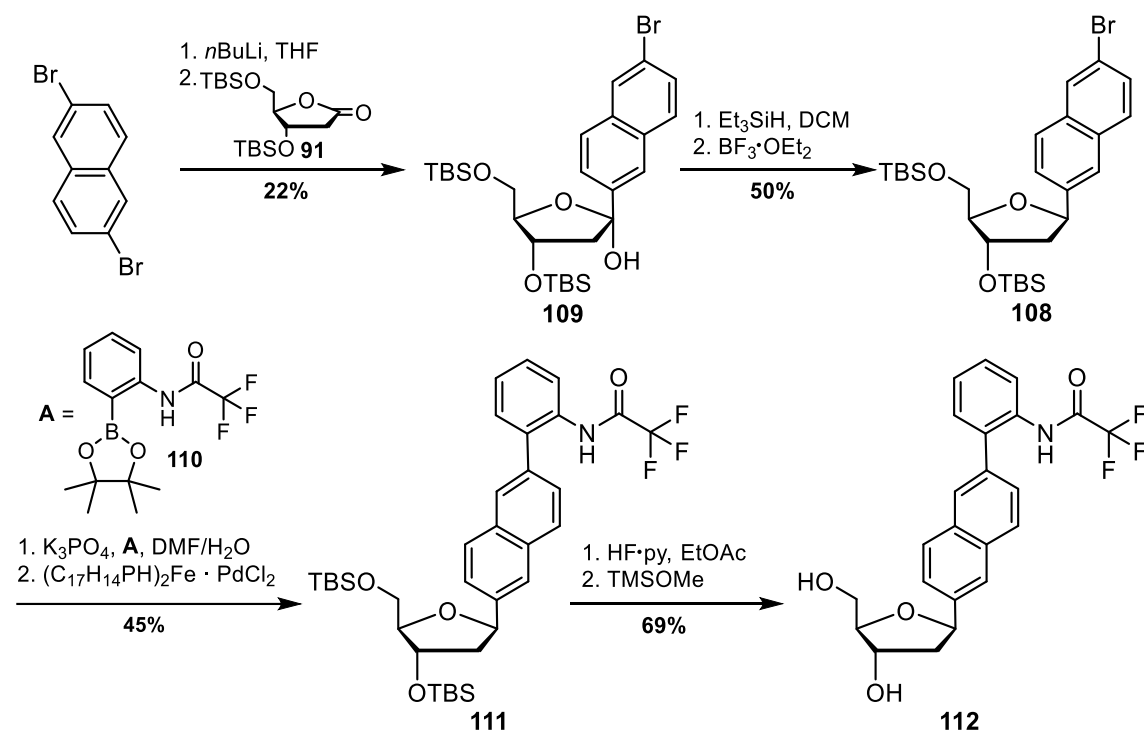


**Figure 3-21.** Assumed preferred conformation of the T<sub>o</sub> base in solution (A), observed conformation in a DNA strand from a X-ray structure (B). Suggested indene- (C) and naphthalene (D) analogs that represent “locked” isosteres of the T<sub>o</sub> base. R = deoxyribose.

#### 3.11.1 Synthesis of a Naphthalene Amine Base Triphosphate

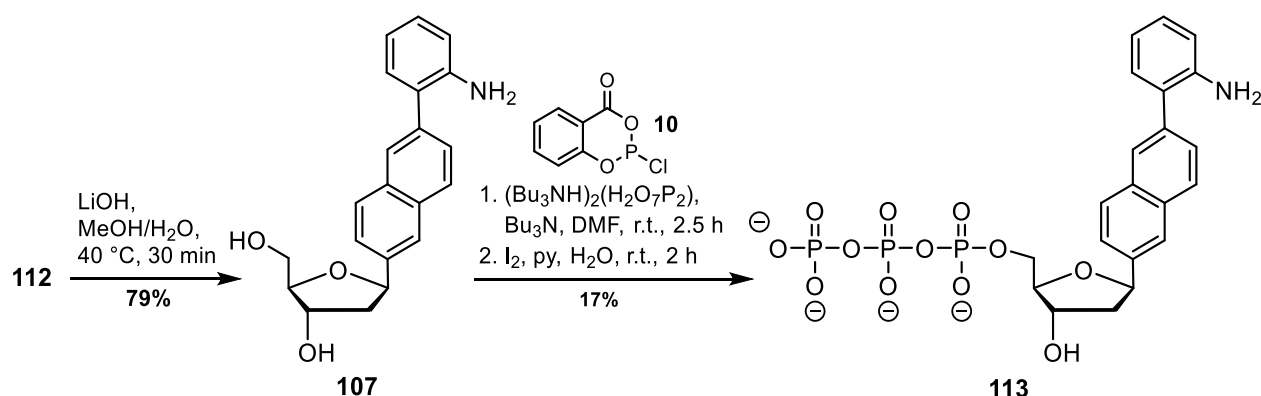
The synthesis of the naphthalene amine nucleoside **107** was based on the developed modular synthesis that had been used for the amine nucleoside library with one major exception. After C-deoxynucleoside formation *via* the hemiketal and subsequent reduction with Et<sub>3</sub>SiH, a Suzuki-Miyaura reaction was employed instead of the Heck coupling. Only minor or no adjustments were necessary for the subsequent reaction steps.

The reaction sequence starts by addition of the monolithiated form of 2,6-dibromo-naphthalene to the TBS-protected deoxyribolactone **91** (Scheme 3-16) analogously to the procedure that was applied in the synthesis of the other amine bases. However, when the crude reaction product from this carbonyl addition was directly used in the BF<sub>3</sub> catalyzed reduction, it was impossible to completely purify the resulting TBS-protected bromo-naphthalene nucleoside **108** by chromatography after the reaction. Therefore, the hemiketal **109** had to be isolated by flash column chromatography before the subsequent reduction to obtain the pure TBS-protected bromo-naphthalene nucleoside **108**. As the naphthalene hemiketal **109** was more stable compared to the benzene hemiketal, this only resulted in minor loss of the product. Again, the formation of an  $\alpha$ -configured bromo-naphthalene nucleoside was not observed.



**Scheme 3-16.** Overview of the TFA-protected naphthalene amide nucleoside **112** synthesis.

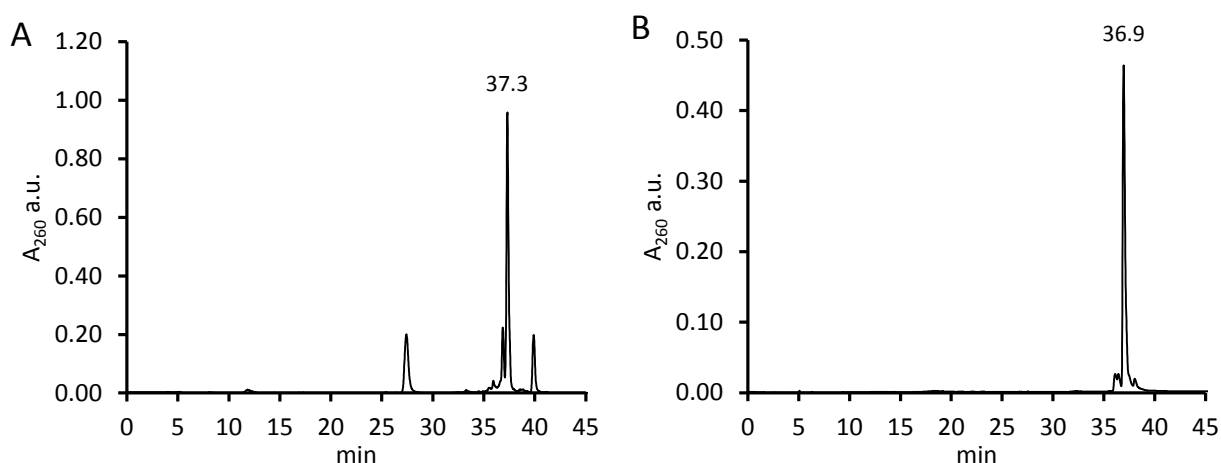
In the Suzuki-Miyaura cross-coupling the TFA-protected boronate **110** was coupled to the bromonaphthalene nucleoside **108** using standard conditions and the ferrocene Pd(II) catalyst,  $\text{Pd}(\text{dppf})\text{Cl}_2$  (Scheme 3-16). Although the starting materials were not completely dissolved, the desired cross-coupling product **111** was formed in 45% isolated yield. Side products were either TFA or onefold TBS-deprotected forms of the coupling product, which were recycled in subsequent reactions to the unprotected nucleoside. For the TBS-deprotection standard conditions involving HF·pyridine were used, which gave the TFA-protected naphthalene amine nucleoside **112** in 69% isolated yield. As the enzymatic acceptance was in the main focus of research at that time, no phosphoramidite was synthesized from the nucleoside **112**.



**Scheme 3-17.** Synthesis of the naphthalene amine triphosphate **113** (dNaaTP) from TFA-protected nucleoside **112**.

Saponification of the TFA-amine nucleoside **112** using mild conditions with LiOH gave the completely unprotected naphthalene amine nucleoside **107** in 79% yield after flash column chromatography (Scheme

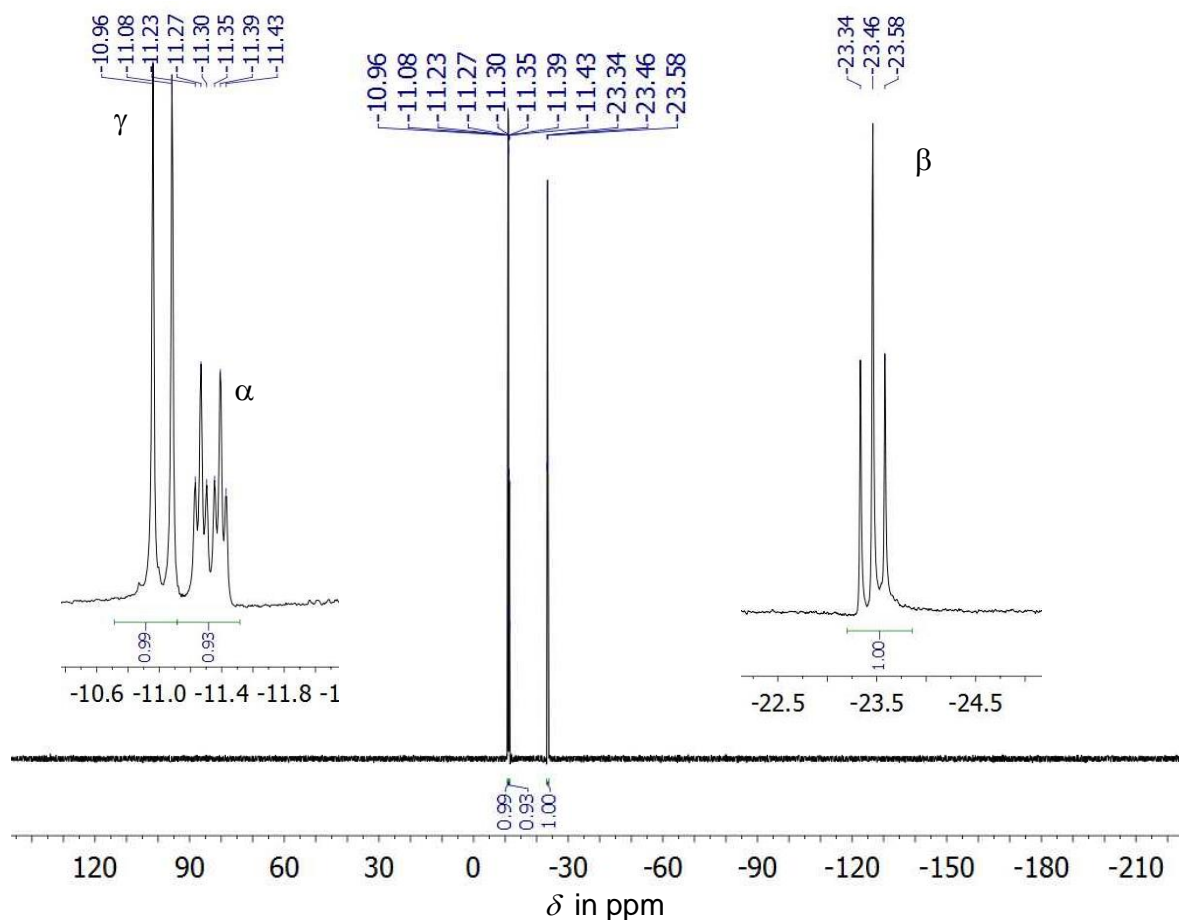
3-17). The reaction to the 5'-triphosphate of the naphthalene amine deoxynucleoside (dNaaTP) **113** was again performed using the self-synthesized salicyl phosphite reagent **10** according to the procedure from *Caton-Williams et al.* which had been successfully applied to several other nucleosides before in this work. After precipitation, the crude product was purified twice by semipreparative RP-HPLC and the resulting fractions were analyzed by MALDI-TOF and analytical RP-HPLC. Satisfyingly, NMR spectroscopy and mass spectrometry of the major combined fraction revealed that 5'-dNaaTP was successfully obtained in 17% isolated yield. The base peak in the analytical RP-HPLC profile from the crude product (Figure 3-22, A) was found to be the desired triphosphate when compared to the purified dNaaTP (Figure 3-22, B). Depending on the number of triethylammonium counterions of the triphosphate small shifts (up to 1 min) in retention time were observed, which lead to an underestimation of the purity. According to analytical RP-HPLC, two-fold purified dNaaTP only had a purity of about 90%. However, judged by the  $^{31}\text{P}$  and  $^1\text{H}$  NMR results at least 95% pure 5'-triphosphate was obtained.



**Figure 3-22.** Analytical RP-HPLC profiles from (A) crude dNaaTP precipitated after the triphosphate reaction and (B) purified dNaaTP at 260 nm detection using a gradient from 0–50% in 45 min.

To ensure the correct formation of the 5'-triphosphate,  $^1\text{H}$ -coupled  $^{31}\text{P}$  NMR experiments were performed. Three signals were observed for the dNaaTP in the  $^{31}\text{P}$  NMR spectrum, of which two only showed  $^2J_{\text{P-P}}$  couplings of about 20 Hz and one also  $^3J_{\text{P-H}}$  couplings of about 6 Hz (Figure 3-23). The latter signal was assigned to the  $\alpha$ -phosphate and due to the multiplet, a doublet of triplets, it was clearly identified as being attached to a 5'-position. The signal at  $-23.5$  ppm couples with two phosphorus and therefore the resulting triplet was assigned to the  $\beta$ -phosphate. For the remaining phosphorus from the  $\gamma$ -phosphate the expected doublet was observed at  $-11.0$  ppm.

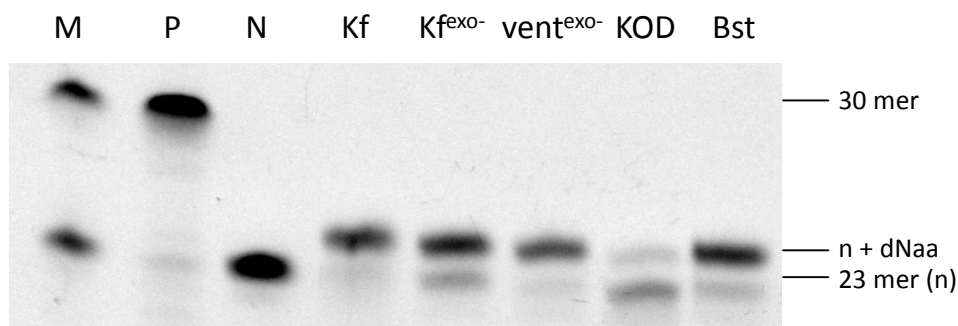
Only minor amounts of the 3'-triphosphate were observed (data not shown) in other fractions from the semipreparative RP-HPLC, which demonstrated the decent chemo- (amine versus hydroxyl nucleophile, primary versus secondary alcohol) and regioselectivity (5'- versus 3'-OH) of the reaction method.



**Figure 3-23.**  $^{31}\text{P}$  NMR spectrum of the purified dNaaTP. The three signals were assigned to the  $\gamma$ -,  $\alpha$ - and  $\beta$ -phosphate. The doublet of triplets for the  $\alpha$ -phosphate proves the formation of the 5'-triphosphate.

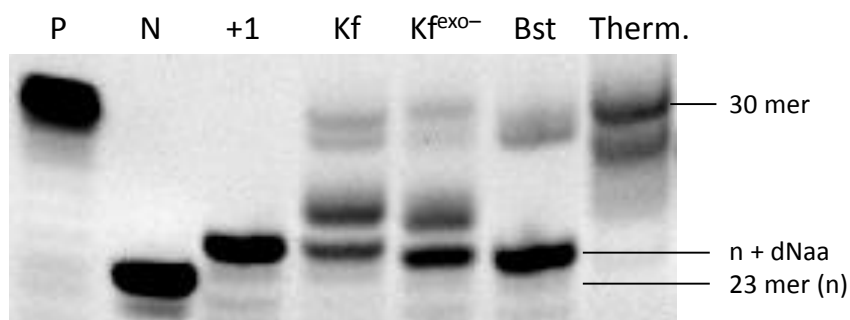
### 3.11.2 Primer Extension Experiments with dNaaTP

With the dNaaTP in hand and the DNA template still available from the experiments with dT<sub>o</sub>TP, the acceptance of the new amine base by the polymerases was studied. At first, single nucleotide insertions of dNaaTP opposite the S using different polymerases and 200  $\mu\text{M}$  of the nucleotide were performed. Within 10 min of incubation at the adequate temperature for each polymerase quantitative incorporation was observed for KF, vent exo<sup>-</sup> and *Bst* Pol I. For KF<sup>exo-</sup> almost complete insertion, and for KOD XL only very little insertion was found (Figure 3-24). Compared to the results obtained in similar experiments for dT<sub>o</sub>TP this already indicated an improvement.



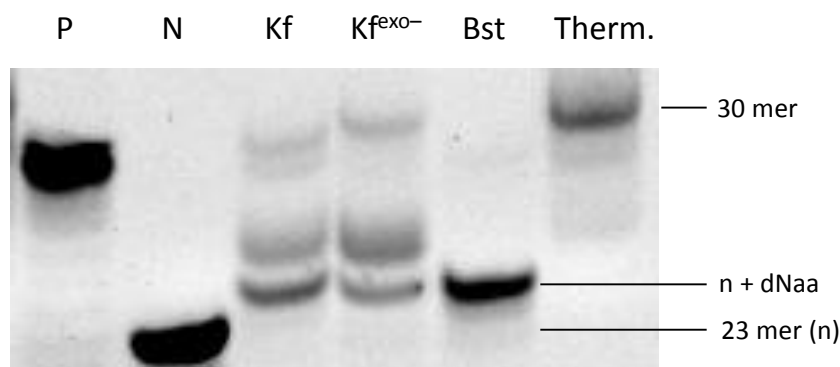
**Figure 3-24.** Denaturing PAGE from SNI experiments in the presence of 200  $\mu$ M dNaaTP using different polymerases. M is the marker, P is the positive control and N is the negative control. Polymerases: KF = Klenow fragment, KF<sup>exo-</sup> = Klenow fragment exonuclease deficient, vent<sup>exo-</sup> = Deep vent exonuclease deficient, KOD = KOD XL, Bst = *Bst* Pol I

Encouraged by this result, a full-length extension was attempted by applying a two-step protocol. In the first step the dNaaTP is incorporated in a SNI using KF for 1 min followed by addition of all canonical dNTPs and assaying different polymerases for 30 min incubation after heat inactivation of KF. At the beginning several polymerases (KF, KF<sup>exo-</sup>, vent<sup>exo-</sup>, KOD XL, *Bst* Pol I, *Taq*, Q5 and Terminator) were tested in this way. However, with the exception of the Terminator polymerase and to a minor extent KF no elongation beyond the +1 was observed (data not shown). So, for a consecutive study these two polymerases and in addition *Bst* Pol I and KF<sup>exo-</sup> were selected for a two-step elongation protocol with an increased incubation time (3 h) for the second step. This time, fully elongated primer was observed for all four polymerases (Figure 3-25). However, KF, KF<sup>exo-</sup> and *Bst* Pol I majorly produced the +1 elongated primer and only traces of the full-length product. Besides for the Klenow fragment polymerases, considerable amounts of +2 were found, indicating that the subsequent elongation after natural nucleotide incorporation adjacent to the dNaa is difficult. For the Terminator polymerase no primer was visible. Instead elongated primer products of 30 or almost 30 mer size were found, indicating reasonable acceptance of the dNaaTP nucleotide.



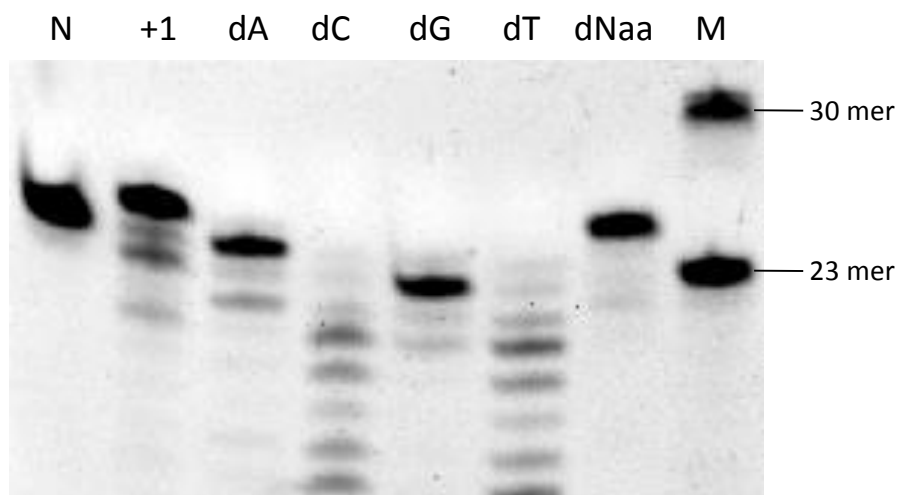
**Figure 3-25.** Denaturing PAGE from a two-step primer extension experiment. First SNI using KF and 200  $\mu$ M dNaaTP, then 200  $\mu$ M dNTPs and different polymerases. P is the positive control, N is the negative control and +1 is the first step from the two-step protocol (3 h incubation)

In parallel, the same polymerases were tested for their ability to fully elongate the primer in the presence of all five dXTPs from the beginning in a one-step protocol. To our delight, the Therminator polymerase was able to produce the fully elongated primer, whereas KF and KF<sup>exo-</sup> only produced tiny amounts and majorly +1 and further stalled products. *Bst* Pol I however exclusively produced the +1 product (Figure 3-26).

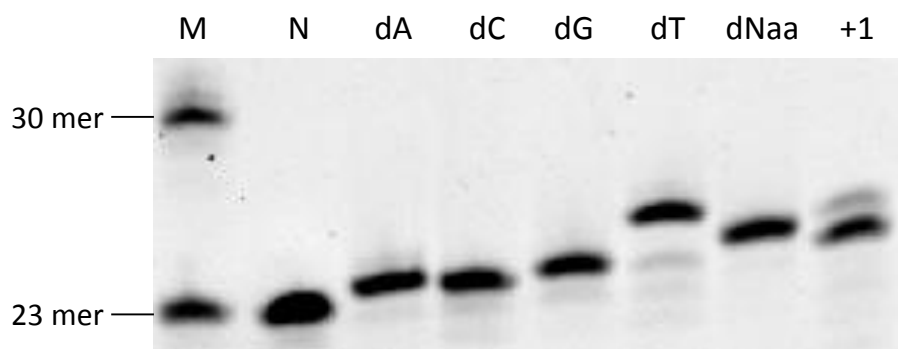


**Figure 3-26.** Denaturing PAGE from a one-step primer extension experiment using different polymerases and 200  $\mu$ M of the 5 dXTPs. Polymerases: KF = Klenow fragment, KF<sup>exo-</sup> = Klenow fragment exonuclease deficient, Bst = *Bst* Pol I, Therm. = Therminator.

Having found a polymerase which was able to fully elongate the primer after incorporation of the dNaaTP, it was necessary to prove the correct replication and to exclude a simple mutation at the artificial base position. Therefore, primer extension experiments were conducted to determine the selectivity for the incorporation of the dNaaTP opposite the templating S by the KF and Therminator polymerases. As the KF had been used for the experiments with S:T<sub>o</sub> and in the two-step protocol for the +1 step for dNaaTP, a SNI was performed using each dXTP. No misincorporation was observed for any of the canonical dNTPs. While the dNaaTP exclusively yielded the +1 elongated primer, no incorporation was observed for any of the canonical dNTPs. Instead, the primer was degraded until the last correct position of the nucleotide which was applied to the primer extension reaction (Figure 3-27).<sup>[193-194]</sup> When the same experiment was performed with the Therminator polymerase, which was the best in the one-step elongations, a completely different result was obtained. This time +1 elongation was observed for dA, dC, dG, and dNaaTP. In the case of dTTP even +3 elongation was found. Under the conditions used the Therminator polymerase simply incorporates any dXTP opposite the templating S, yet it only continues elongation beyond the mismatch, if there are matching base pairs like for dT (Figure 3-28). This result illustrated the ability of the polymerase to tolerate mismatches and incorporate modified nucleotides. Whether meaningful elongation of the primer without mutation was possible in the presence of the 5 dXTPs therefore remained questionable.



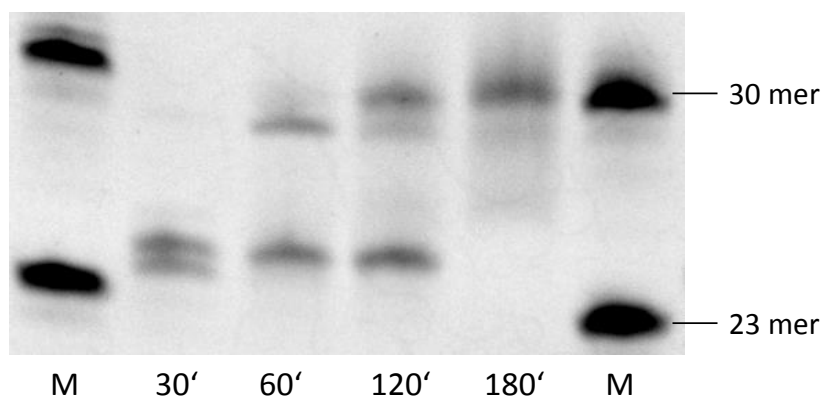
**Figure 3-27.** Denaturing PAGE from a SNI using KF polymerase and 200  $\mu$ M of each dXTP. M is the marker, N is the negative control and +1 is the singly elongated control.



**Figure 3-28.** Denaturing PAGE from a SNI using the Terminator polymerase and 200  $\mu$ M of each dXTP. M is the marker, N is the negative control and +1 is the singly elongated control.

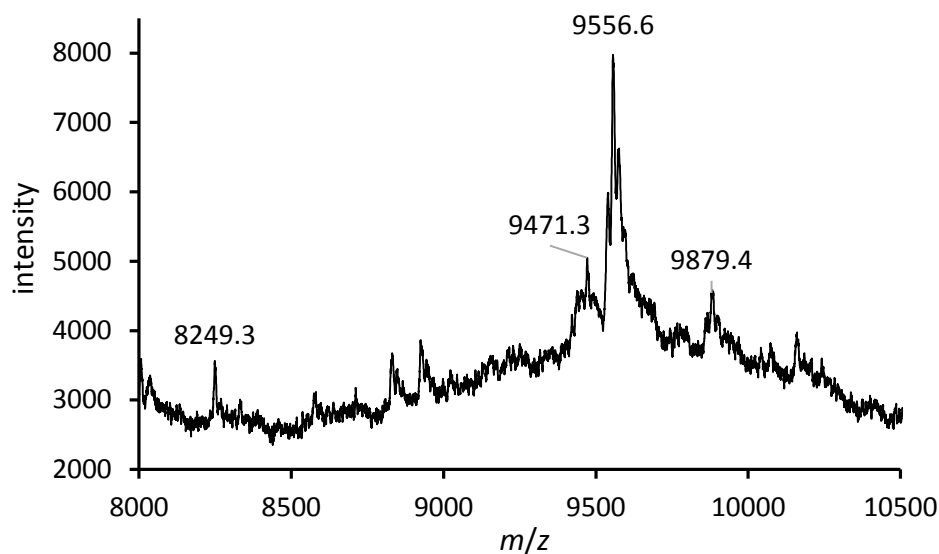
In order to clarify this issue and to study the time-dependence of the full elongation, primer extension reactions involving the Terminator polymerase and all 5 dXTPs were stopped after 30, 60, 120 and 180 min. A sample of a 180 min extension was used for MALDI-TOF mass spectrometry after phenol-chloroform extraction. Within 30 min the Terminator polymerase only manages to incorporate a single base opposite the templating S, after 60 min some further elongated primer is visible which is remarkably shorter than the full-length product from the marker (Figure 3-29). Only after 120 min some fully elongated primer occurs, yet the presumed +1 product is clearly visible. Complete consumption is observed for the 3 h sample, but the resulting band is accompanied by a smear, which has also been observed for primer extensions involving dT<sub>o</sub>TP.





**Figure 3-29.** Denaturing PAGE from primer extensions using the Terminator polymerase and 200  $\mu$ M of each dXTP for different incubation times.

Finally, the MALDI measurement of the 180 min sample clarified the specificity issue (Figure 3-30). It was known that dA has the highest probability of misincorporation opposite a templating S,<sup>[13]</sup> so the mass spectrum was searched for the corresponding strands (Table 3-5). The main peak was assigned to a +6 elongated primer possessing a single dNaa with the found  $m/z$  of 9556.6 in agreement with the calculated mass (9558) and is probably the main product from the primer extension (Figure 3-30). This means that the final dA incorporation at the very end of the template did not occur for this signal (Table 3-5, entry 5). For the full-length product including the dNaa a  $m/z$  of 9879.2 (+7) was found, which deviates from the expected 9873 considerably. However the low signal to noise ratio for this peak might attribute for the mass difference and also indicates a low abundancy of the full-length product. Yet, it remains unclear why the incorporation of the canonical dNTPs towards the end of the template is so difficult in the case of the dNaa incorporation to the primer. Peaks which indicate dA misincorporation opposite the templating S could be the  $m/z$  of 8249.3 and 9471.3, which are relatively close to the 8240 and 9474 that are expected for the primer elongated by AT instead of NaaT and ATTGCT instead of NaaTTGCT (Table 3-5, entry 2 and 6).



**Figure 3-30.** MALDI-TOF spectrum of primer extension products from one-pot reactions with 200  $\mu\text{M}$  dNaaTP, dNTPs and 20 pmol of the salicylaldehyde containing template using one unit of Therminator polymerase for 3 h at 70  $^{\circ}\text{C}$ .

**Table 3-5.** Masses of possibly elongated primers from the primer extension experiment.

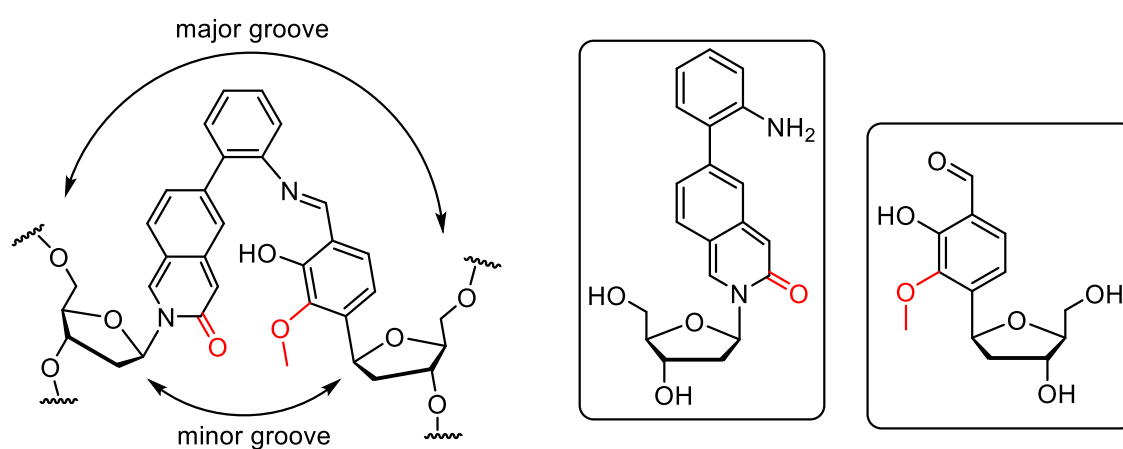
Entry	Primer +	Expected mass X = A, g/mol	Expected mass X = Naa, g/mol
1	X	7934	8018
2	XT	8240	8324
3	XTT	8546	8630
4	XTTG	8877	8961
5	XTTGC	9168	9252
6	XTTGCT	9474	9558
7	XTTGCTA	9789	9873

More surprisingly, the MALDI measurement from the one-pot primer extension hints towards a remarkable selective incorporation of the dNaaTP opposite the dS, when all dXTPs are present. This is in contrast to the result from the primer extension experiment aiming at determining the selectivity of the Therminator polymerase in SNIs, which showed possible incorporation of all dNTPs opposite of dS, when no dNaaTP was present (Figure 3-28). Altogether this is evidence for the kinetically favored acceptance of the dNaaTP by the Therminator polymerase opposite dS compared to the natural dNTPs, as it does not possess a 3'-5' exonuclease activity.

In conclusion, it was shown that the enzymatic acceptance of the amine base could be improved by a minor chemical redesign of the initial  $T_0$  base. The synthesis of the naphthalene amine base triphosphate, dNaaTP, was accomplished in only six steps from commercially available starting materials. Primer extension experiments proved that full elongation and selective incorporation of dNaaTP opposite the dS is possible in the presence of all dXTPs, when using the Therminator polymerase. This is a tremendous

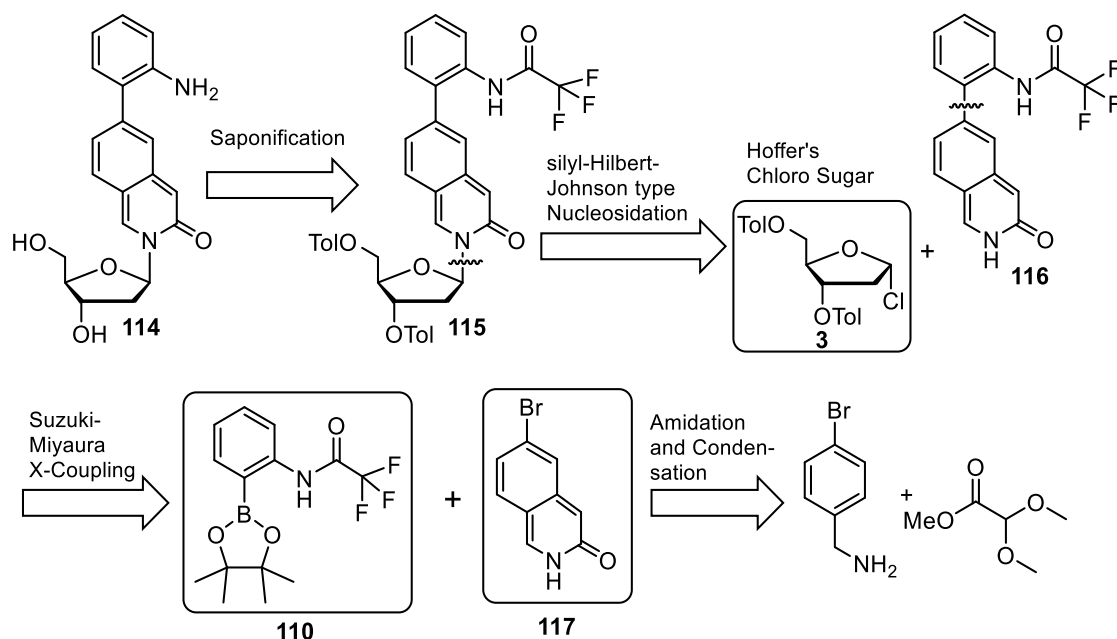
improvement compared to the  $T_0$  base, which did not allow efficient full elongation even when polymerase mixtures or a sequential order of component addition was used. Reasons for the improved acceptance of the dNaa in comparison to the d $T_0$  base might be the increased  $\pi$ -stacking ability and the stronger rigidity. By introducing the additional cyclic system, the degree of conformational freedom is reduced, and thus the loss of entropy upon binding and adaption of a defined conformation is decreased. Ironically, the acceptance by *Bst* Pol I which was in the X-ray structure together with the initial S: $T_0$  base pair and allowed the deduction of the new amine base design, still halts after the dNaaTP incorporation, indicating a major problem with the “improved amine base”.

### 3.11.3 Design and Proposed Synthesis of Further Improved Covalent Base Pairs



**Figure 3-31.** Proposed design of the further improved covalent base pair and the corresponding nucleosides. The H-bond accepting functional group is marked in red.

The existence of a minor groove H-bond acceptor is one striking common property of published artificial base pairs which feature enzymatic acceptance by polymerases similar to the canonical bases regarding efficiency of the incorporation rate and selectivity.<sup>[135, 163, 195-196]</sup> The X-ray crystal structure from the Klen taq polymerase in a ternary complex with a DNA containing the dNaM base in the template and an incoming d5SICS triphosphate showed close contacts of a tyrosine residue and the minor groove H-bond acceptor.<sup>[197]</sup> This finding is in accordance with existing structures also for natural DNA<sup>[198]</sup> and seems to be a determining factor for DNA synthesis processivity by the polymerase. The mechanism adds up to the other control factors, like  $\pi$ -stacking ability and correct geometry of the newly formed base pair. Therefore, the logical consequence for the further improvement of the developed covalent base pair from our group was the introduction of a suitable functional group capable of accepting H-bonds. At the same time the synthetic effort was to be kept at a minimum.



**Scheme 3-18.** Retrosynthetic analysis of the isoquinoline amine nucleoside. The molecule can be traced back to three major building blocks, of which two are already known.

In order to prove the hypothesis that the H-bond accepting moiety will improve the acceptance by the DNA polymerase, synthesis of the 5'-triphosphate isoquinoline nucleoside and evaluation in primer extension studies against the initial dS base, was planned. The isoquinoline amine base nucleoside **114** could be accessed in a straightforward 5-step reaction sequence (Scheme 3-18). A final saponification of fully protected isoquinoline amine nucleoside **115** might give the desired nucleoside **114**. Before the deprotection, protected molecule **115** could be provided *via* a silyl-Hilbert-Johnson type nucleosidation from Hoffer's chlorosugar **3** and TFA-protected isoquinoline amine **116**. Analogously to the synthesis of the naphthalene amine nucleoside, a Suzuki-Miyaura cross-coupling between TFA-boronate **110** and bromo-isoquinoline derivative **117** is intended for the synthesis of base building block **116**. The synthesis of boronate **110** was already established during the synthesis of naphthalene amine nucleoside **107** (chapter 3.11.1, p. 68). For the bromo-isoquinoline **117** an amidation condensation reaction from 4-bromobenzamine and methyl dimethoxyacetate was proposed similar to existing procedures.<sup>[199-200]</sup>

## Part III – Click Chemistry Labeling of Phosphate-modified Nucleotides

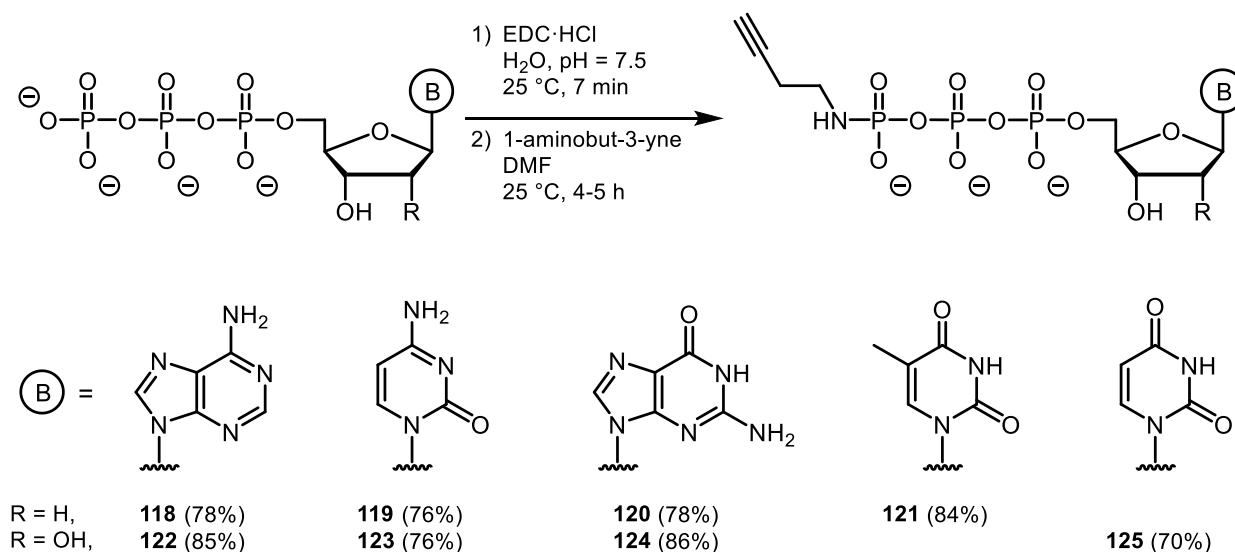
Phosphates are important functional units in signal transduction pathways,<sup>[46, 201]</sup> which regulate protein-protein interactions<sup>[202]</sup> and enzyme activity.<sup>[203]</sup> Due to their charge they considerably change the hydrophobicity of the modified protein, which can lead to a local conformation change and ultimately to an altered protein structure.<sup>[204]</sup> Most of the organic phosphate in our body is captured in the form of the energy providing molecule ATP. At any given day, every one of us turns over an ATP amount equivalent to their body weight.<sup>[45]</sup> Also, DNA and RNA is synthesized from nucleotides within our body and the phosphate backbone of the DNA provides the necessary connectivity and charge.

Therefore, new methods to create modified phosphate groups is a promising field for future applications. Many of the existing phosphate modifications were introduced by solid-phase synthesis and were chemically addressed with the Cu(I)-catalyzed alkyne-azide cycloaddition reaction. By this means, a DNA with a clickable backbone was generated that allowed to chemically ligate a linear DNA to a circular form.<sup>[98]</sup> Also the biocompatibility of the generated triazole backbone was demonstrated by *in vivo* transcription of a red fluorescent protein in human cells.<sup>[101]</sup> Some enzymatic approaches rely on the natural promiscuity of the enzymes to incorporate substrate analogs to a certain extent and thereby introduce alkynes or azides for subsequent modification *in vitro*.<sup>[205-206]</sup> This method has the potential to generate long site-specifically modified RNA strands, which hardly can be obtained by solid-phase synthesis. Another obstacle of this approach is the chemical synthesis of the modified substrate itself, which often is long and tedious. This part summarizes efforts about the synthesis, incorporation and application of alkyne-phosphate modified nucleotides.

### 3.12 Synthesis of $\gamma$ -Labeled Nucleotides

Despite of immense technical challenges, studies involving single enzyme molecules are in vogue.<sup>[102]</sup> These developments are fueled by the advancement of the technical instrumentation for analysis and the ability to create smaller devices with defined properties. At the same time the design and synthesis of reporter molecules that generate a signal upon enzymatic conversion, is an ongoing challenge. In order to improve the crucial signal to noise ratio, these reporters are usually fluorescent probes mimicking substrates or cofactors of the applied enzyme of interest.<sup>[207]</sup> Unwanted alterations of the enzyme activity are minimized by introducing these modifications in positions that barely interact with the enzyme. Even real-time information can be acquired without the need for special miniature technical equipment. For example, ubiquitin activation has been studied with an ATP reporter molecule, which possesses a FRET donor-acceptor pair.<sup>[14]</sup> Single molecule real-time (SMRT) sequencing represents a combination of these technologies. In principle, this sequencing technique not only provides sequence information of a single DNA molecule, but also information about naturally occurring modified bases like 5-methyl-, 5-hydroxymethyl-, 5-formyl- and 5-carboxy-deoxycytidine *via* altered incorporation kinetics.<sup>[3, 208-210]</sup> This

method requires  $\gamma$ -phosphate modified deoxynucleotides with a different fluorophore on every nucleotide. Upon addition to the growing primer strand, the fluorophore is released concomitantly with pyrophosphate.<sup>[3]</sup> Known syntheses of  $\gamma$ -labeled nucleotides involve multistep procedures with low overall yields and require anhydrous conditions, which are unfavorable for the water-soluble triphosphates.<sup>[7, 211]</sup> Consequently, there is potential for improvement in the synthesis of  $\gamma$ -labeled nucleotides. Our idea was to introduce an alkyne linker at the  $\gamma$ -phosphate of the nucleotide for subsequent modification by click chemistry.<sup>[212]</sup>

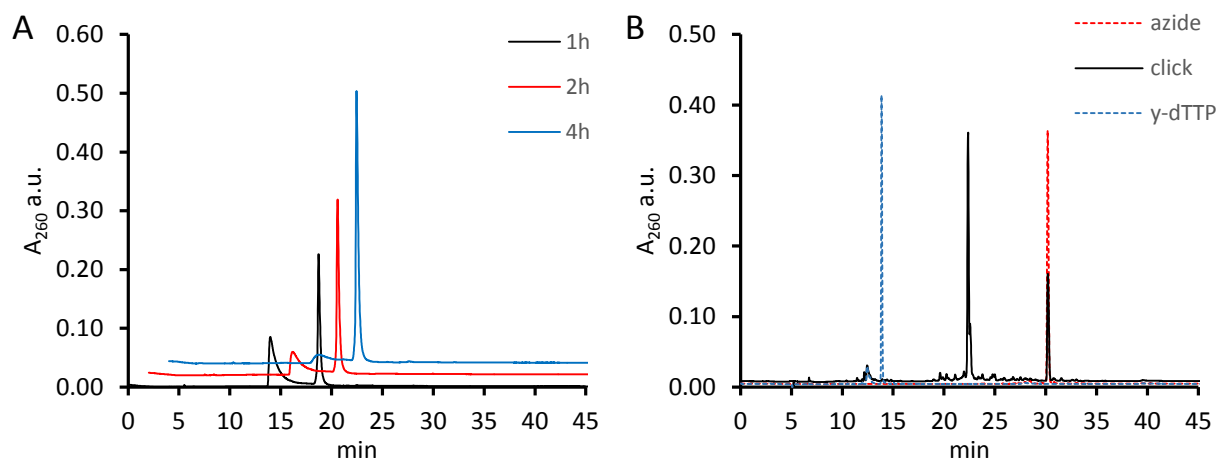


**Scheme 3-19.** Synthesis of alkyne labeled nucleotides from dNTP and NTP sodium salts. Note that the products **118-125** were isolated as their tris- and tetrakis triethylammonium salts after RP-HPLC.

Based on the conditions from *Morr et al.* 1-ethyl-3-(3-dimethyl-aminopropyl) carbodiimide hydrochloride (EDC·HCl) was used as a coupling reagent, in order to attach 1-aminobut-3-yne at the  $\gamma$ -phosphate of the triphosphates.<sup>[213]</sup> This afforded the alkyne  $\gamma$ -phosphoramidates in good isolated yields of 70% to 86% for all major eight natural nucleotides (Scheme 3-19). Other coupling reagents like carbonylimidazole and *N,N'*-dicyclohexylcarbodiimide afforded lower yields in accordance with the literature and required anhydrous conditions.<sup>[214]</sup> Screening for best reaction conditions, different coupling reagents and alternative synthesis strategies was performed by *F. M. Kink* during his master's thesis.<sup>[215]</sup>

Despite the high functional group density of nucleotides, no protecting groups were necessary to selectively obtain the  $\gamma$ -phosphate alkyne triphosphates by our method. The correct configuration was unambiguously proven by NMR. Indicative is the coupling between the  $\alpha$ -phosphorus and 5'-H in <sup>31</sup>P-<sup>1</sup>H HMBC measurements (data not shown). Moreover, analytical RP-HPLC of aliquots from different time points of the reaction allowed tracking of the reaction progress. As the reaction time progresses, the deoxycytidine triphosphate amount decreases, while the  $\gamma$ -alkyne deoxycytidine triphosphate **119** amount increases (Figure 3-32, A). Apart from the starting material and the product peak, no additional peak is observed in the chromatogram, thus demonstrating the selectivity of the coupling method. In our final

protocol, the solvent DMF and the 1-aminobut-3-yne were extracted with  $\text{CHCl}_3$  and the remaining aqueous layer was used for NaCl-ethanol precipitation of the  $\gamma$ -modified triphosphate. Thereby the purity of the “crude” product was improved considerably and possible side reactions during concentration of the aqueous layer were avoided.



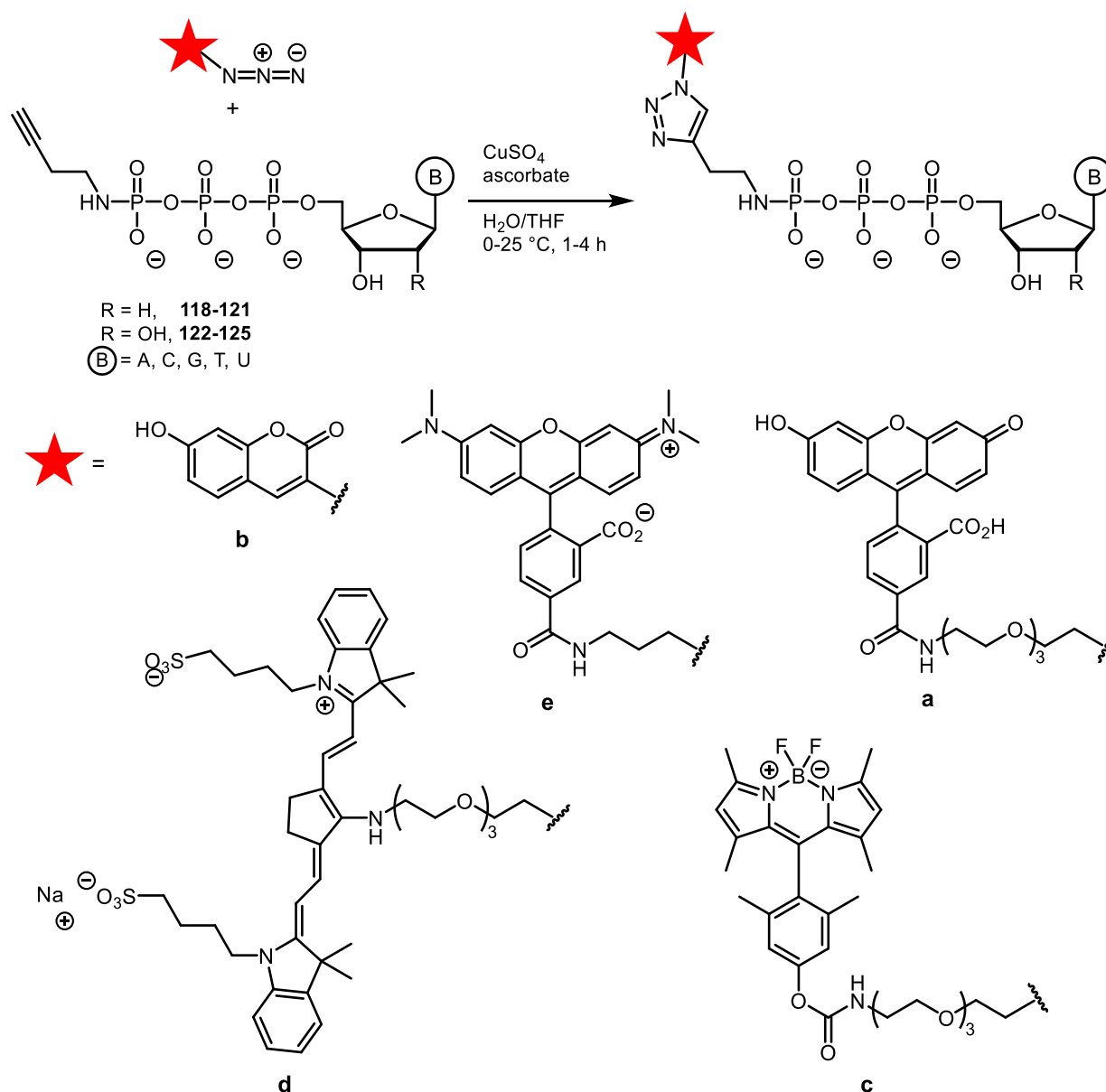
**Figure 3-32.** HPLC profile of (A) aliquots from the  $\gamma$ -alkyne modification of dCTP and (B) crude product of  $\gamma$ -fluorescein dTTP after 1 h reaction time (black). (B) The starting materials,  $\gamma$ -alkyne dTTP **121** (blue dotted) and the fluorescein azide **a** (red dotted) are depicted for comparison.

Subsequently, Cu(I)-catalyzed alkyne-azide cycloaddition (CuAAC) was used to provide a set of fluorophore modified nucleotides.<sup>[71-72]</sup> In order to find the suitable conditions for the CuAAC between the alkyne triphosphates and the fluorophore azides, we took advantage of a coumarin azide **b** (Scheme 3-20). This azide is non-fluorescent, yet becomes a fluorophore upon reaction to the triazole. Moreover, it can be easily accessed in a few steps from commercially available starting materials.<sup>[216]</sup> The initial “click-conditions” screening between the  $\gamma$ -dTTP alkyne **121** and the coumarin azide **b** was performed by *F. M. Kink*.<sup>[215]</sup> Reaction success was evaluated *via* the fluorescence intensity upon irradiation with a UV-lamp for thin-layer-chromatography analysis at 366 nm and analytical RP-HPLC. The final protocol involves *in situ* formation of Cu(I) from  $\text{CuSO}_4$  and fresh sodium ascorbate in a THF/ $\text{H}_2\text{O}$  mixture. Only a slight excess of the fluorophore azide over the alkyne triphosphate was used at 0-25 °C. In this way the  $\gamma$ -coumarin dTTP **121b** and  $\gamma$ -coumarin ATP **122b** were synthesized in 70% and 77% isolated yield, respectively. Characterization by HRESIMS and NMR spectroscopy allowed to verify product formation. Encouraged by the result, fluorophore azides with different absorbance and emission maxima were used for the click reaction with the  $\gamma$ -alkyne nucleotides. Since the fluorophores, which are routinely applied in SMRT sequencing are expensive,<sup>[7]</sup> additional fluorophore azides, namely BODIPY **c**, carboxyfluorescein **a**, and an analog of an alexa fluorophore **d** were synthesized (Scheme 3-20). Two of the used fluorophores, the carboxyfluorescein and the alexa analog IR-806 are commercially available and were simply transformed to an azide by attaching a PEG-azide-amine by amide coupling or conjugate substitution, respectively. The BODIPY fluorophore was generously provided as a phenolic alcohol by *Dr. B. Hackner*.

Briefly, the phenolic alcohol was reacted to a succinimidyl carbonate which was *in situ* converted to BODIPY **c** by addition of the PEG-azide-amine linker. Due to chemical and physical properties of the PEG-azide, the introduction of the linker-azide was facile and the hydrophilicity of the resulting fluorophore azides was increased.

The small set of fluorophore azides was used in click reactions of  $\gamma$ -dTTP **121** under the condition established for the coumarin azide **b**. This condition was ideal for the carboxyfluorescein azide **a** and provided the triazole product within 1 h in at least 90% conversion yield as shown by analytical HPLC (Figure 3-32, B). However, for the BODIPY azide **c** almost no product was formed using these conditions. Instead, precipitation of the fluorophore azide **c** was observed, which indicated a major solubility problem for the fluorophore in the 1:1 mixture of THF/H<sub>2</sub>O. By increasing the THF content, the solubility of the BODIPY azide **c** was improved. Still the reaction progress was unsatisfactory, as most of the starting material remained after 4 h. Using additional amounts of CuSO<sub>4</sub> and ascorbate however, almost quantitative conversion ( $\geq 90\%$ ) was finally achieved for the reaction between  $\gamma$ -dTTP **121** and BODIPY azide **c** within 3 h. Despite special precautions, which avoided light exposure of fluorophore azide **d**, the resulting click product and the azide **d** itself decomposed within a few hours in solution. The decomposition was accompanied by a change in color from blue to violet.





**Scheme 3-20.** Click chemistry based synthesis of  $\gamma$ -fluorophore labeled nucleotides from  $\gamma$ -alkyne labeled nucleotides **118-125** and fluorophore azides **a-e** (red star).

In order to complete the set of four different fluorophore cycloadducts, additional fluorophore azides were purchased. The limit of the click protocol was illustrated by the poor yields, which were obtained with the Eterneon 480/635 azide<sup>®</sup> as a click partner (data not shown). Due to the poor solubility of the azide in solvent mixtures that contain enough water to establish sufficient  $\text{CuSO}_4$  and ascorbate concentrations, only trace amounts of the “clicked” nucleotide were isolated. In contrast to this, the water soluble TAMRA azide **e** allowed quantitative conversion of the  $\gamma$ -alkyne nucleotide within 2 h. Probably, a click protocol applying mostly hydrophilic fluorophore azides is convenient for nucleotides. It is important to note that no  $\text{Cu(I)}$ -stabilizing ligands like TBTA or THPTA were necessary to achieve full conversion in 1–4 h in the presence of only 1.5 equivalents of fluorophore azide per alkyne nucleotide. Inspired by the high purity of the crude  $\gamma$ -alkyne products and the subsequent click reaction, a one-pot protocol for the  $\gamma$ -coumarin dTTP

**121b** was developed. After formation of the  $\gamma$ -alkyne dTTP **121**, the crude product was directly applied to the click reaction with coumarin azide **b**. In this way, the  $\gamma$ -coumarin dTTP **121b** was obtained in 60% yield from dTTP after HPLC purification. Thereby, one time-consuming preparative HPLC step was avoided and the yield was identical compared to the calculated total yield of the step-wise procedure.

Altogether, four different  $\gamma$ -fluorophore labeled dTTPs **121a-c**, **121e** were synthesized by CuAAC in 70-90% yield (Table 3-6, Entry 4 – 7). The fluorophores possessed different absorbance and emission maxima, as evident from their physical data and visual inspection at daylight and upon excitation at 366 nm (Figure 3-33). Chemically different dyes (rhodamine, fluorescein, coumarin and BODIPY) illustrate that the protocol is mild and tolerates a variety of functional groups. In addition,  $\gamma$ -fluorescein nucleotides were prepared from the eight  $\gamma$ -alkyne nucleotides (Table 3-6, Entry 1 – 4, 9 – 12). No considerable difference in yield and reaction progress was observed between the nucleotides by analytical RP-HPLC.

**Table 3-6.** Overview of the successful CuAAC reactions.

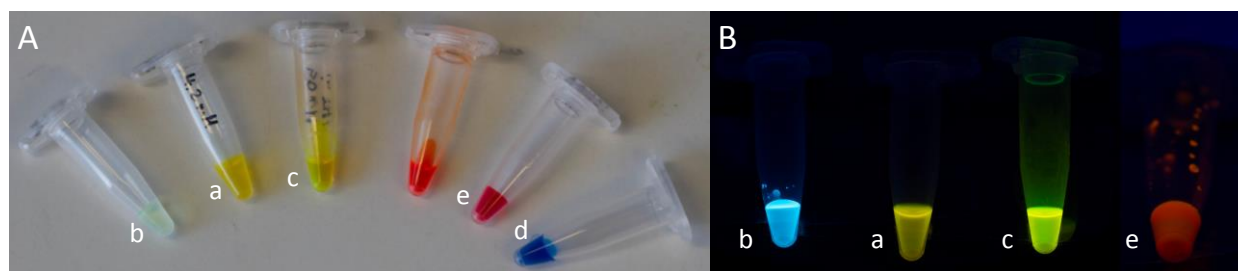
entry	$\gamma$ -NTP	azide	product	conditions	yield (%) <sup>1)</sup>	chemical formula [M – H] <sup>–</sup>	MS calc. m/z	MS found m/z
1	dATP	<b>a</b>	<b>118a</b>	A	≥90	C <sub>43</sub> H <sub>48</sub> N <sub>10</sub> O <sub>20</sub> P <sub>3</sub> <sup>–</sup>	1117.2265	1117.2221
2	dCTP	<b>a</b>	<b>119a</b>	A	≥90	C <sub>42</sub> H <sub>48</sub> N <sub>8</sub> O <sub>21</sub> P <sub>3</sub> <sup>–</sup>	1093.2152	1093.2110
3	dGTP	<b>a</b>	<b>120a</b>	A	≥90	C <sub>43</sub> H <sub>48</sub> N <sub>10</sub> O <sub>21</sub> P <sub>3</sub> <sup>–</sup>	1133.2214	1133.2168
4	dTTP	<b>a</b>	<b>121a</b>	A	≥90	C <sub>43</sub> H <sub>49</sub> N <sub>7</sub> O <sub>22</sub> P <sub>3</sub> <sup>–</sup>	1108.2149	1108.2113
5	dTTP	<b>b</b>	<b>121b</b>	A	≥90, 70 <sup>2)</sup>	C <sub>23</sub> H <sub>26</sub> N <sub>6</sub> O <sub>16</sub> P <sub>3</sub> <sup>–</sup>	735.0624	735.0620
6	dTTP	<b>c</b>	<b>121c</b>	C	≥90	C <sub>44</sub> H <sub>60</sub> BF <sub>2</sub> N <sub>9</sub> O <sub>18</sub> P <sub>3</sub> <sup>–</sup>	1144.3336	1144.3344
7	dTTP	<b>e</b>	<b>121e</b>	B	≥90	C <sub>42</sub> H <sub>49</sub> N <sub>9</sub> O <sub>17</sub> P <sub>3</sub> <sup>–</sup>	1044.2465	1044.2438
8	ATP	<b>b</b>	<b>122b</b>	A	≥90, 77 <sup>2)</sup>	C <sub>23</sub> H <sub>25</sub> N <sub>9</sub> O <sub>15</sub> P <sub>3</sub> <sup>–</sup>	760.0688	760.0687
9	ATP	<b>a</b>	<b>122a</b>	A	≥90	C <sub>43</sub> H <sub>48</sub> N <sub>10</sub> O <sub>21</sub> P <sub>3</sub> <sup>–</sup>	1133.2214	1133.2224
10	CTP	<b>a</b>	<b>123a</b>	A	≥90	C <sub>42</sub> H <sub>48</sub> N <sub>8</sub> O <sub>22</sub> P <sub>3</sub> <sup>–</sup>	1109.2101	1109.2122
11	GTP	<b>a</b>	<b>124a</b>	A	≥90	C <sub>43</sub> H <sub>48</sub> N <sub>10</sub> O <sub>22</sub> P <sub>3</sub> <sup>–</sup>	1149.2122	1149.2163
12	UTP	<b>a</b>	<b>125a</b>	A	≥90	C <sub>42</sub> H <sub>47</sub> N <sub>7</sub> O <sub>23</sub> P <sub>3</sub> <sup>–</sup>	1110.1942	1110.1913

1) yield determined by analytical RP-HPLC; 2) isolated yield after RP-HPLC purification; reaction conditions:

A) 1.5 eq fluorophore azide, 0.1 eq CuSO<sub>4</sub>, 2.0 eq sodium ascorbate, 0 °C to 25 °C, 1 h, THF/H<sub>2</sub>O = 1:1;

B) 1.5 eq fluorophore azide, 0.1 eq CuSO<sub>4</sub>, 2.0 eq sodium ascorbate, 0 °C to 25 °C, 2 h, THF/H<sub>2</sub>O = 1:3.

C) 1.5 eq fluorophore azide, 0.3 eq CuSO<sub>4</sub>, 6.0 eq sodium ascorbate, 25 °C, 3 h, THF/H<sub>2</sub>O = 3:1.



**Figure 3-33.** (A) Optical appearance of the fluorophore-labeled dTTP **121a-121e** and the Eterneon 480/635 azide<sup>®</sup> (red orange) at daylight. (B) Fluorophore-labeled dTTP **121a-c, 121e** upon excitation with light at 366 nm.

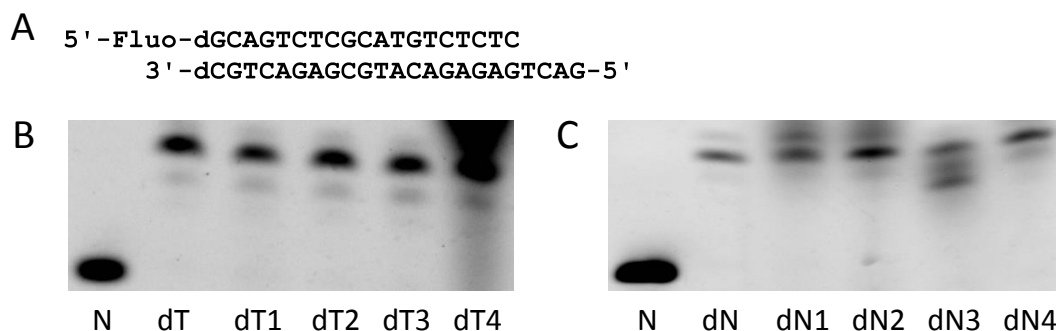
As phosphoramidates are prone to hydrolysis, especially under acidic conditions,<sup>[217]</sup> we studied the stability of the  $\gamma$ -alkyne **121** and  $\gamma$ -fluorescein dTTP **121a** at pH 4, 7 and 10 over 3 h at 37 °C or 72 °C. For comparison, natural dTTP was incubated under the same conditions as well. All samples were analyzed by RP-HPLC after incubation. At pH 4, the  $\gamma$ -alkyne dTTP **121** was completely hydrolyzed to dTTP, while about 75% of the  $\gamma$ -fluorescein dTTP **121a** was still intact. No hydrolysis was observed for any of the modified nucleotides at pH 7. Interestingly, at pH 10 the dTTP was less stable than its modified congeners. While for the natural triphosphate up to three peaks were observed, after incubation at 72 °C, about 98% of the  $\gamma$ -fluorescein dTTP **121a** was still unaltered. Since most of the enzymes, which are routinely used in molecular biology require a neutral pH, we concluded that the chemical stability of our modified nucleotides is sufficient.

### 3.13 Enzymatic Incorporation of $\gamma$ -Labeled Nucleotides into DNA

Next, the use of  $\gamma$ -labeled nucleotides as substrates for nucleotide consuming enzymes, especially DNA polymerases, was investigated. Therefore, primer extensions were performed involving the Klenow fragment (KF) from *E. coli*. The template was designed in a way that for the full extension of the primer each dNTP had to be incorporated exactly once. During incorporation of the fluorophore nucleotides, the label is cleaved off together with pyrophosphate resulting in a natural DNA strand (without any additional modification). To ensure an unambiguous result of the primer extension experiments, all applied labeled nucleotides were carefully purified by RP-HPLC.

In one set of experiments the natural dTTP was completely replaced by one of the four different  $\gamma$ -fluorophore labeled dTTPs **121a-c, 121e**. The samples were analyzed by denaturing PAGE. A fluorescein on the 5'-position of the primer (Figure 3-34, A) allowed observation of the assay products *via* their fluorescence without staining. For all four  $\gamma$ -fluorophore dTTPs, complete primer extension was observed (Figure 3-34, B). However, the incorporation of fluorescein-labeled dTTP **121a** seemed problematic, as an additional band was observed (Figure 3-34, B, dT4). In the second set of experiments, the influence of the same label (fluorescein) on the incorporation of different deoxynucleotides was studied. Therefore, one of the four dNTPs was replaced by the  $\gamma$ -fluorescein congener in consecutive experiments. Full extension was

observed for almost all samples in the denaturing PAGE. Again, for the fluorescein-labeled dTTP **121a** an additional band was observed (Figure 3-34, C, dN3), which indicates compatibility issues for this substrate.



**Figure 3-34.** (A) Template strand for the primer extension experiments. (B) Denaturing PAGE from primer extension experiments replacing dTTP by  $\gamma$ -fluorophore dTTPs (1: BODIPY, 2: coumarin, 3: TAMRA, 4: fluorescein) and (C) replacing a single dNTP each by a  $\gamma$ -fluorescein dNTP (1: dA\*, 2: dG\*, 3: dT\*, 4: dC\*).

### 3.14 Enzymatic Incorporation of $\gamma$ -Labeled Nucleotides into RNA

We next studied the enzymatic compatibility of the  $\gamma$ -modified ribonucleotides. To this end, the  $\gamma$ -alkyne and  $\gamma$ -fluorophore ribonucleotides were used for *in vitro* transcription experiments. In procaryotic transcription, the first added ribonucleotide remains as a triphosphate, thus generating 5'-triphosphate RNA oligonucleotide transcripts. By replacing this nucleotide with a  $\gamma$ -labeled nucleotide, the modification could be introduced into the RNA strand during transcription. Moreover, the labeling is site-specific, as the modification of any subsequently introduced  $\gamma$ -labeled nucleotide is cleaved off together with pyrophosphate. Similar efforts, applying an  $\alpha$ -dienophile or  $\alpha$ -azide modified monophosphate, have produced RNA transcripts that can undergo inverse electron demand Diels-Alder reactions with tetrazines or Cu(I) catalyzed alkyne/azide cycloadditions, respectively.<sup>[205, 218]</sup>

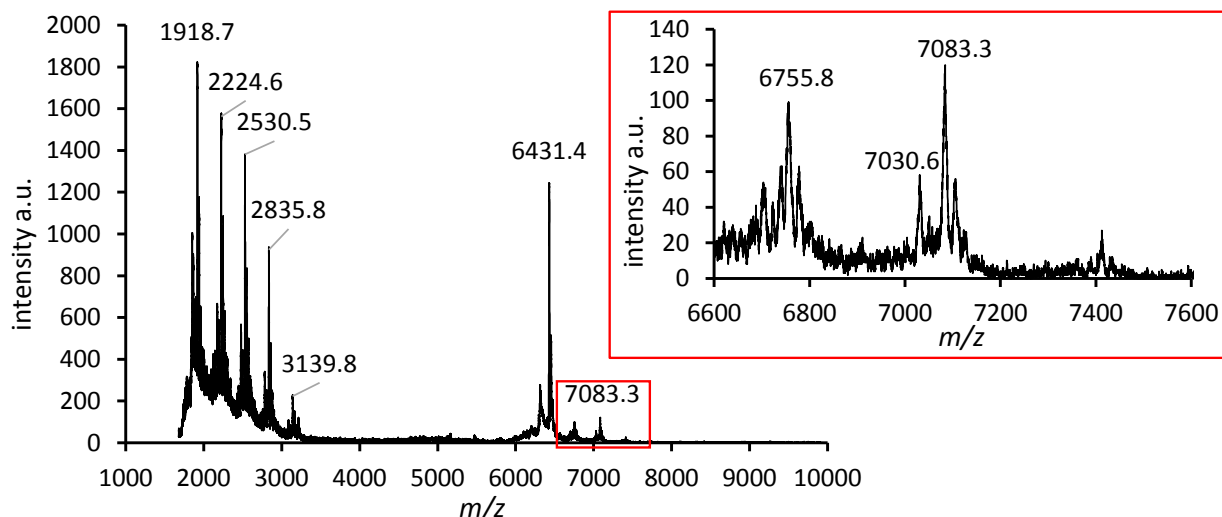
The DNA template for the transcription was designed as an early-stage T7 RNA polymerase promotor and a single stranded region coding for a 21-mer transcript. Since most of the T7 promotors have a transcription start site that begins with the incorporation of GTP, the canonical GTP was replaced by  $\gamma$ -labeled GTP in the transcription experiments. The *in vitro* transcription outcome was analyzed by MALDI-TOF mass spectrometry. Different T7 RNA polymerase mutants were studied for their ability to accept the labeled nucleotides, as they were available from former transcription experiments (see 3.5 Transcription Experiments using T7 RNA Polymerase Mutants, p. 43).

**Table 3-7.** Templates used for *in vitro* transcription experiments and resulting full-length transcripts including calculated mass.\*5'-alkyne label, <sup>1</sup>non-templated A addition, <sup>2</sup>5'-fluorescein label.

Entry	Template and Resulting Transcript	Transcript Mass [M-H] <sup>-</sup>	Transcript* Mass [M-H] <sup>-</sup>
1	5'-ATAATACGACTCACTATAGGG 3'-TATTATGCTGAGTGATATCCCTTAGGGCTCTTCACTGAT  pppGGGAAUCCCGAGAAGUGACUA	7033.9	7085.0
2	5'-ATAATACGACTCACTATAGGCCTTTCCTACTCTCCTACCT 3'-TATTATGCTGAGTGATATCCGGAAAGTGATGAGGATGGA  pppGGCCUUUCACUACUCCUACCU pppGGCCUUUCACUACUCCUACCUA <sup>1</sup>	6741.8 7070.9	6792.9 (7369.3) <sup>2</sup> 7122.0

First experiments with  $\gamma$ -alkyne GTP **124** and wild-type enzyme resulted in formation of alkyne labeled 5-9 mer polyguanosine RNA fragments. When equal concentrations of the four NTPs were added together with the  $\gamma$ -alkyne GTP **124** in the transcription, almost no peaks from alkyne modified fragments were observed. Transcription experiments in which the natural GTP was completely replaced by its alkyne congener, provided a peak at  $m/z$  7083 indicating the successful production of the alkyne-labeled RNA 21-mer transcript (Figure 3-35). As the transcript contains 7 guanosines (Table 3-7, Entry 1), this result shows that the  $\gamma$ -alkyne GTP incorporation in internal positions of the transcript is possible. During this process, the label was released together with pyrophosphate. The intensity of the full-length peak increased when the wild-type RNA polymerase was replaced by the mutant forms F644A and H784A, which are known to possess a less strict substrate selection.<sup>[187]</sup> Also, a delayed addition of the natural GTP after 1 h transcription time considerably improved the yield of the alkyne full-length product. Probably, before the addition, the transcription complexes were progressing slowly during incorporation of  $\gamma$ -labeled GTP at internal positions. Due to the preference of the RNA polymerase for its natural substrate, the GTP replaced the labeled form upon addition and improved the transcription progress. When the time point for the natural GTP addition was delayed to 10 or 30 min after transcription start with  $\gamma$ -alkyne GTP, ATP, CTP and UTP, the yield of unlabeled transcripts increased.

Despite all efforts to optimize the amount of labeled RNA transcripts, the MALDI-TOF mass spectrum was dominated by 5'-alkyne labeled abortive initiation fragments 5-9 mers ( $m/z$  1918.9 to 3140.8, Figure 3-35). Besides this, the mass of the sense strand of the template was observed at  $m/z$  6431.4. Although peaks in the MALDI-TOF mass spectrum are not quantitative, the result indicates the presence of only low amounts of the labeled full-length transcript.

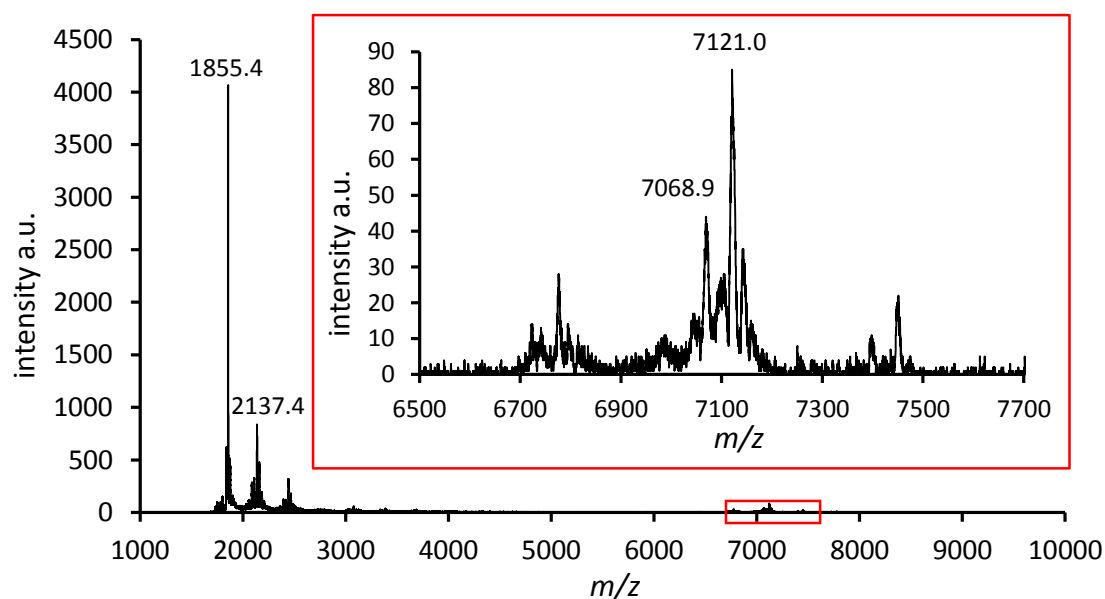


**Figure 3-35.** MALDI-TOF mass spectrum from *in vitro* transcription labeling experiments using wild-type T7 RNA polymerase, 2  $\mu\text{M}$  of template (**Table 3-7**, p. 87, entry 1), 400  $\mu\text{M}$   $\gamma$ -alkyne GTP **124**, ATP, CTP, UTP and after 55 min GTP.

Small interfering RNAs (siRNA) can regulate the posttranscriptional gene expression.<sup>[219]</sup> Several efforts are under way to develop siRNA-based therapeutics, but cell-specific and intracellular delivery of the polyanionic molecules is a major problem.<sup>[220]</sup> Recently, an anandamide receptor-mediated uptake of dendrimeric siRNA was developed within the *Carell* group, which allows targeted delivery of siRNAs to neuronal cells.<sup>[94]</sup> The anandamide siRNA dendrimers were prepared through CuAAC from alkyne-modified RNA and an anandamide-azide branching molecule. All alkyne-modified RNAs were prepared by solid-phase synthesis.<sup>[94]</sup>

Enzymatic synthesis could provide an alternative access to an alkyne containing siRNA. A template which encodes for a part of the luciferase gene from *Renilla reniformis* was employed for *in vitro* transcriptions with  $\gamma$ -alkyne GTP **124**. The resulting transcript could act as a small interfering RNA (siRNA) in a dual-luciferase reporter assay. The reporter assay was used to quantify the cellular uptake of oligonucleotides attached to a dendritic anandamide ligand.<sup>[94]</sup>

As the transcript only contains two guanosines at the start (**Table 3-7**, Entry 2), more facile labeling of the transcript was expected. Again several conditions were screened to optimize the labeling efficiency and the samples were analyzed by MALDI-TOF mass spectrometry. This time, only abortive initiation fragments of 5-7 mers were observed, indicating an improved processivity for the template under the applied conditions (**Figure 3-36**). The signals at  $m/z$  2137 and 2443 were assigned to the alkyne-labeled 6 mer and 7 mer, respectively. Unexpectedly, the signal at  $m/z$  1855, was not just the alkyne labeled 5 mer, but probably contained an additional C to A mutation, as its calculated mass (1859) fits best to the observed result. Nevertheless, the peak for the labeled full-length transcript (additional A,  $m/z$  7121) was small and also the unlabeled form was found ( $m/z$  7069).

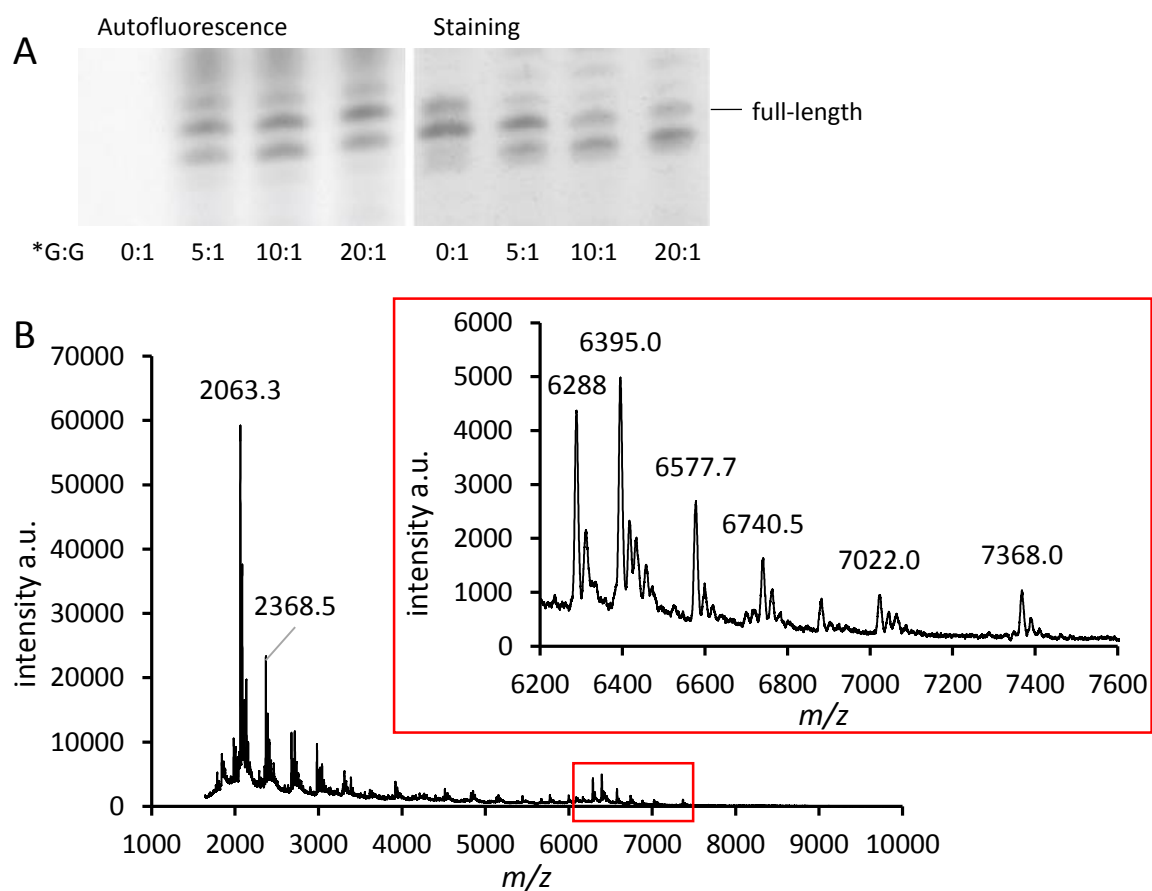


**Figure 3-36.** MALDI-TOF mass spectrum from *in vitro* transcription labeling experiments using wild-type T7 RNA polymerase, 2  $\mu$ M of template (Table 3-7, p. 87, entry 2), 400  $\mu$ M  $\gamma$ -alkyne GTP, ATP, CTP, UTP and after 75 min GTP.

Then, efforts were undertaken to introduce a fluorophore azide to the alkyne-labeled transcript by click chemistry. After extensive dialysis of the alkyne-labeled transcripts, the azidocoumarin **b** (Scheme 3-20, p. 83),  $\text{CuSO}_4$  and ascorbate were added in a THF/ $\text{H}_2\text{O}$  mixture. After 3-4 h the sample started to fluoresce upon excitation at 366 nm, which indicated successful reaction progress. However, only small peaks of the expected product were observed by MALDI-TOF (data not shown). Alternative reaction conditions, which were applied for effective click-labeling of DNA in the *Carell* group,<sup>[91, 93]</sup> could not improve the yield. Isolation of the labeled RNA by RP-HPLC was not possible.

Therefore, direct labeling during *in vitro* transcription using  $\gamma$ -fluorescein GTP **124a** was studied for the Ren *luc* template (Table 3-7, p. 87, Entry 2). At the beginning, the natural GTP was replaced by the fluorescent analog **124a** and after different transcription incubation times (0, 5, 30, 60, 120 min) additional GTP was added. Notably, an early time point of the addition (0 and 5 min) improved the yield of fluorescein labeled transcripts (data not shown). Consequently, all five nucleotides were added to the transcription from the beginning and various concentration ratios of GTP: $\gamma$ -fluorescein GTP were analyzed. A ten- to twentyfold excess of the labeled GTP was found to give the best yields of fluorescein modified transcripts as judged by MALDI-TOF. When the samples were applied to denaturing PAGE, direct visualization of the labeled transcription products was possible *via* their intrinsic fluorescence (Figure 3-37, A). Subsequently, the same gel was treated with SYBR green II staining to visualize non-labeled transcripts as well. Clearly, fluorescein-labeled full-length product was obtained especially when the natural GTP amount was low. This was also confirmed by MALDI-TOF mass spectrometry (Figure 3-37, B), as a peak at  $m/z$  7368.0 was found which is close to the calculated molecular mass of 7369.3. However, the amount of abortive initiation fragments ( $m/z$  2063.3 to 3000) increased enormously as

compared to the control transcription. In addition, peaks corresponding to labeled transcripts lacking several bases were observed from 6000-7000 Da. Despite the successful 5'-fluorescein labeling of the transcript, these findings indicate a significant decrease in transcription efficiency upon incorporation of the  $\gamma$ -fluorescein GTP.



**Figure 3-37.** (A) PAGE and (B) MALDI-TOF mass spectrum from *in vitro* transcription labeling experiments. Conditions: T7 RNA polymerase, 2  $\mu$ M of template (**Table 3-7**, p. 87, entry 2), 400  $\mu$ M  $\gamma$ -fluorescein GTP **124a**, ATP, CTP, UTP and 20-80  $\mu$ M GTP. \*G:G is the ratio of labeled to unlabeled GTP.

### 3.15 Enzymatic Labeling Efforts Involving $\gamma$ -Labeled Nucleotides

The labeling efforts involving the RNA polymerase had shown that 5'-labeling during transcription is possible, yet rather inefficient. Therefore, a different enzyme was studied for its ability to incorporate  $\gamma$ -phosphate labeled nucleotides.

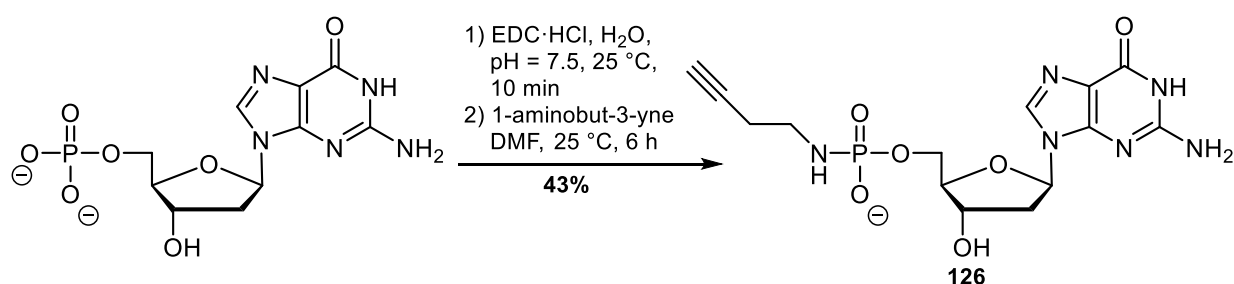
The T4 polynucleotide kinase was examined as it catalyzes the transfer of a  $\gamma$ -phosphate from ATP to the 5'-hydroxyl group of single- and double stranded DNA and RNA.<sup>[221]</sup> With the help of  $\gamma$ -labeled ATPs in the *in vitro* assays, site selective labeling of oligonucleotides was targeted. Using  $\gamma$ -alkyne **122**,  $\gamma$ -fluorescein ATP **122a** and a single stranded DNA oligonucleotide (5'-AGCCAGTACATCACAAGAACTCA-3'), the  $m/z$  equal to the mass of the phosphorylated DNA strand



was observed in the MALDI-TOF mass spectrum. The same result was found in the sample containing ATP, whereas a negative control without ATP did not give any 5'-phosphorylated oligonucleotide. As the employed phosphoramidates are prone to hydrolysis, an adenosine-5'-[ $\gamma$ -(propargyl)]triphosphate was purchased. Phosphate ester hydrolyze much slower compared to the phosphoramidates.<sup>[43]</sup> Again, the usage of this more stable nucleotide resulted in the formation of the 5'-phosphorylated DNA. Obviously, the reaction catalyzed by the T4 polynucleotide kinase not only facilitates the hydrolysis of the triphosphate, but also of other residues attached to the  $\gamma$ -phosphate. Maybe, by providing a  $\gamma$ -alkyne phosphonate ATP, this issue could be solved. However, as other groups have already synthesized  $\gamma$ -azide phosphonate ATP<sup>[217]</sup> and did not report on its application in enzymatic labeling, we assume that it is incompatible with this enzyme.

### 3.16 Synthesis and Application of an $\alpha$ -Alkyne Deoxyguanosine 5'-Triphosphate

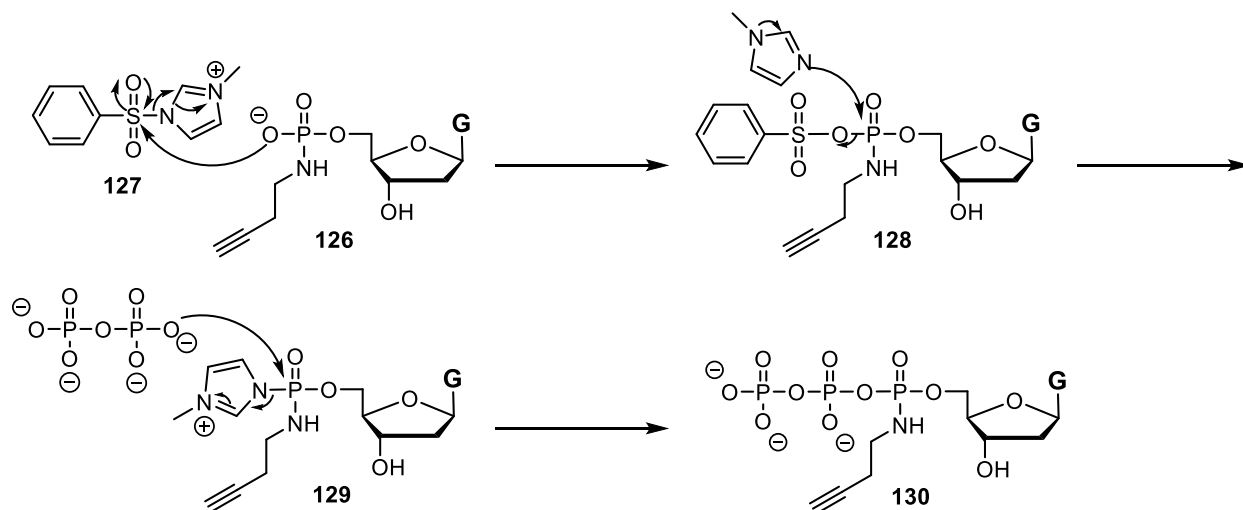
Inspired by the simple synthesis of the  $\gamma$ -labeled nucleotides,  $\alpha$ -labeled nucleotides were next synthesized. If these were consumed during enzymatic DNA synthesis, the introduction of a modified phosphate into the DNA backbone would become possible. The synthesis strategy started by linking the 1-aminobut-3-yne to the nucleoside monophosphate using already established protocols (see 3.12).<sup>[212]</sup> In a subsequent reaction, the labeled monophosphate was converted to the triphosphate following a published procedure.<sup>[53]</sup> EDC-mediated reactions of 1-aminobut-3-yne with the monophosphate of dAMP, dCMP, dGMP and dTMP were analyzed by analytical HPLC. Considerable differences in conversion were observed, which demonstrated the influence of the nucleobase on the reactivity of the 5'-phosphate. As the yield for the reaction to the alkyne-dGMP **126** was most promising, the initial synthesis strategy was established with dGMP first. Slight adaption of the initial protocol furnished the  $\alpha$ -alkyne dGMP **126** in 43% yield after RP-HPLC isolation (Scheme 3-21).



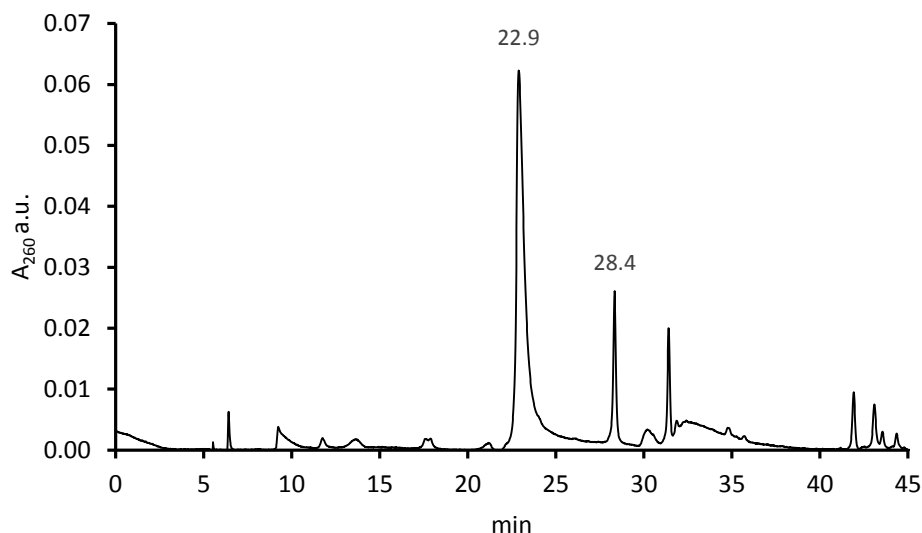
**Scheme 3-21.** Synthesis of the  $\alpha$ -alkyne dGMP **126** from deoxyguanosine monophosphate.

The reaction from the monophosphate to the triphosphate was based on a published procedure, whose basic principle is from the 1970s, and was chosen as it provides the natural triphosphates in at least 80% yield without the need for protecting groups.<sup>[53]</sup> At the beginning the monophosphate is activated under basic conditions and attacks the sulfonylimidazolium salt **127** (Scheme 3-22). The resulting sulfonyl phosphorylanhydride **128** reacts with the imidazole and gives an activated phosphorus species **129**. This

phosphorimidazolidate **129** is *in situ* converted to the triphosphate by attack of pyrophosphate. In order to circumvent that the phosphorimidazolidate **129** is quenched by water and acetate, the starting materials needed to be free of water and triethylammonium acetate buffer (from HPLC purification). Formation of the phosphorimidazolidate **129** can be examined with the help of its yellow color. After addition of excess amounts of pyrophosphate to the activated compound, the yellow color remained even for prolonged incubation (>24 h). This indicated almost no reaction progress, as upon triphosphate formation the color should disappear as the intermediate is consumed. Analytical RP-HPLC confirmed the observation, as most of the starting material was still visible and only a small product peak of the desired  $\alpha$ -alkyne GTP **130** was found (Figure 3-38). The best yield was achieved using 4 equivalents of pyrophosphate and incubation at 25 °C for 8 h. Then, the  $\alpha$ -alkyne GTP was isolated in 2% yield after careful purification by RP-HPLC. This only allowed characterization of the compound by high-resolution ESIMS and UV-Vis. Probably, the additional residue on the phosphate group decreases the electron density, thus rendering the phosphorimidazolidate more stable. Therefore, the attack by the pyrophosphate is less favorable. Also sterical hindrance from the butyne moiety could account for the very low yield of the  $\alpha$ -labeled GTP **130** compared to the reported triphosphate yields from the natural monophosphate.<sup>[53]</sup>



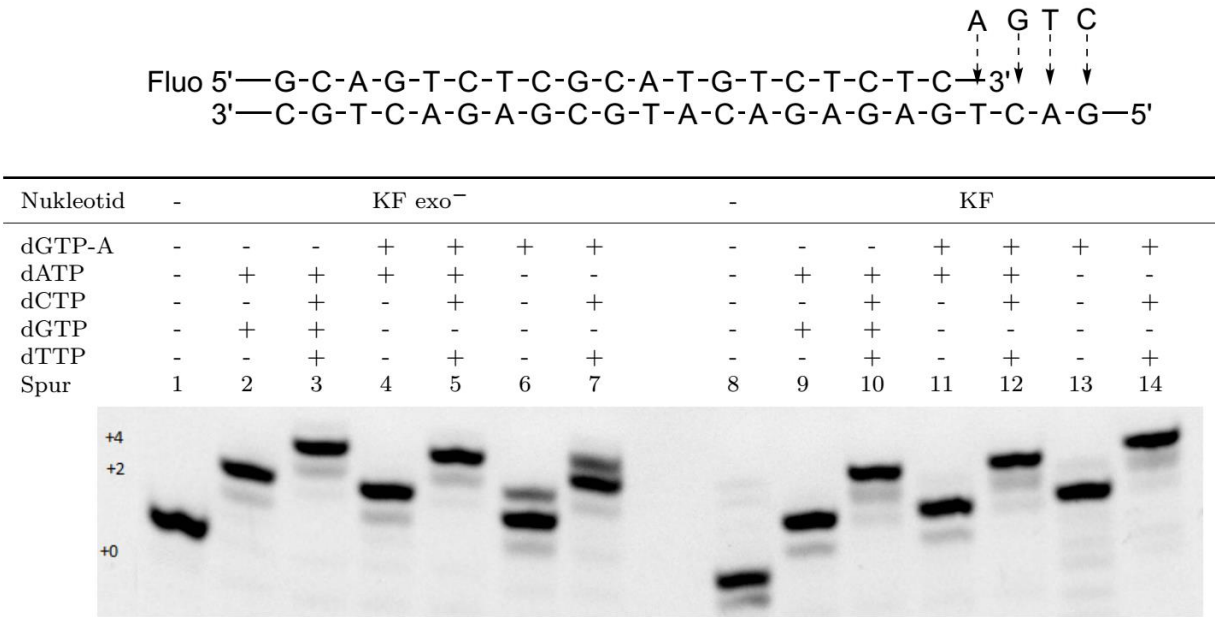
**Scheme 3-22.** Proposed mechanism of the reaction from  $\alpha$ -alkyne dGMP **126** to  $\alpha$ -alkyne dGTP **130** adapted from Mohamady *et al.*<sup>[53]</sup>



**Figure 3-38.** Analytical HPLC profile from the reaction of  $\alpha$ -alkyne dGMP **126** (22.9 min) to  $\alpha$ -alkyne dGTP **130** (28.4 min) after 8 h reaction time and detection at 260 nm.

Despite of the poor yield of  $\alpha$ -alkyne dGTP **130**, the amounts were sufficient to perform primer extension experiments in which the modified triphosphate **130** replaced natural dGTP. The results provided insight whether the labeled nucleotide is a substrate for the enzymatic introduction of DNA backbone modifications. Therefore, a template strand with each of the 4 major natural bases missing (19 mer hybrid template) and a template ending directly before dGTP incorporation (20 mer hybrid template) was employed. These were then incubated together with an exonuclease-deficient and an exonuclease active DNA polymerase and varying nucleotide compositions. The reaction outcome was analyzed by denaturing PAGE.

For almost all samples, in which the natural dGTP was replaced by the  $\alpha$ -alkyne dGTP **130** (Figure 3-39, Lane 4-7,11-14), the same bands were observed as for the positive controls (Figure 3-39, Lane 2-3,9-10). Only for two samples, using the exonuclease-deficient Klenow polymerase and the 20 mer hybrid template, incomplete elongation was found (Figure 3-39, Lane 6-7). However, this finding cannot be explained by inefficient incorporation of the labeled dGTP, as the almost identical experiment with the 19 mer hybrid template yielded the +2 and +4 elongated primer (Figure 3-39, Lane 4-5).



**Figure 3-39.** Primer extension experiments involving  $\alpha$ -alkyne dGTP **130**. The template shown here is the 19 mer hybrid template; the corresponding 20 mer hybrid template is elongated by dA. The table summarizes the nucleotide composition of the experiments. Note that in lane 6, 7, 13 and 14 the 20 mer hybrid template was used.

The result that the  $\alpha$ -alkyne dGTP **130** gives efficiently elongated primer like the natural dGTP is unexpected. Particularly for the exonuclease active polymerase elongation stop after incorporation of the labeled nucleotide was anticipated. In addition, no gel-shift for the supposed labeled DNA was observed in comparison to the non-labeled strand. This aroused the suspicion that the label might have disappeared during/before incorporation. Several efforts were undertaken to unambiguously clarify whether enzymatic incorporation of the  $\alpha$ -alkyne dGTP **130** provided indeed a labeled DNA strand. Finally, analytical RP-HPLC of the  $\alpha$ -modified dGTP **130**, which had been used in the primer extension experiment, proved hydrolysis to dGTP (about 5%). Since an excess of the  $\alpha$ -alkyne nucleotide was used in the primer extensions, even the small amount of the natural dGTP generated by hydrolysis was sufficient to provide efficient primer elongation.

From the results, it was obvious that the phosphoramidate based  $\alpha$ -modified nucleotides were by no means suitable substrates to analyze whether enzymatic DNA backbone labeling is possible. Future efforts should rely on labels introduced *via* phosphate esters or even phosphonates, to avoid unwanted hydrolysis. This however, will impede the synthetic effort.

## 4 Conclusions and Outlook

Several artificial base pairs are reported in the literature, which can be replicated by DNA polymerases.<sup>[5, 9-10, 13, 122, 133, 163]</sup> Only a fraction of these base pairs can also be successfully transcribed<sup>[58, 116, 123, 181, 196]</sup> into RNA and subsequently translated into proteins.<sup>[119, 126]</sup> Strategies relying on alternative hydrogen bonding<sup>[58, 118, 120, 126]</sup> or hydrophobic interactions<sup>[5, 8-10, 42, 123, 130, 139, 181]</sup> are commonly used to create unnatural base pairs. Work from within the *Carell* group shows that the enzymatic replication of an artificial metal-base pair is possible with good fidelity.<sup>[13]</sup> Transcription and translation of this base pair was the aim at the beginning of this PhD thesis. All necessary compounds were prepared successfully and combined for *in vitro* transcriptions. Different aldehyde nucleotides and T7 RNA polymerase mutants were studied under a plethora of conditions. Despite all efforts, transcription of the salicylaldehyde base was not possible without mutation of the artificial base to a canonical nucleoside in the resulting transcript.

As a consequence, the base pair strategy was adopted and metal ion complexation was avoided completely. Instead, reversible covalent imine chemistry was exploited as the sole source of interstrand interaction. Therefore, the existing salicylaldehyde interacted with a suitable amine-containing base, which was designed to fit into a DNA duplex. In cooperation with Dr. *M. Tomás-Gamasa*, a small set of amine base phosphoramidites was designed and synthesized. The salicylaldehyde phosphoramidite was utilized in the solid-phase synthesis of an aldehyde containing strand. Likewise, different amine-containing strands were prepared and the interaction between aldehyde- and amine-containing DNA strands was studied in melting temperature experiments. The combination between an aromatic amine and the salicylaldehyde proved to be highly stabilizing for the DNA duplex and the melting temperature increased by up to 31 K, compared to the control duplex. Notably, the melting curves obtained for heating and cooling of this strand were not superimposable and thus displayed hysteresis. By applying different cooling and heating rates or by increasing or decreasing the pH, this hysteresis effect was changed. Based on the data, we concluded that the imine formation and hydrolysis within the duplex is slower than the H-bond mediated melting and annealing of the adjacent bases. The comparison to a benzaldehyde base, which lacked the *ortho*-hydroxylgroup of the salicylaldehyde, clearly proved its importance for stabilizing the imine by hydrogen bonding.

Encouraged by the observed interaction between the aromatic amine and the salicylaldehyde in melting temperature experiments, incorporation of the artificial bases by DNA polymerases was analyzed. Therefore, the corresponding aromatic amine (dT<sub>o</sub>TP) and salicylaldehyde triphosphates (dSTP) were synthesized and applied in primer extension experiments. A combination of Klenow fragment and the polymerase *Bst* Pol I was found capable of faithfully elongating a primer from an aromatic amine containing template, when dSTP and all dNTPs were present from the start. Base pair formation of the dT<sub>o</sub>TP opposite the templating salicylaldehyde was possible but subsequent elongation could not be achieved.

In order to gain insight into the isosteric nature of the base pair, a DNA containing the pair was co-crystallized with *Bst* Pol I. The structure revealed that within the duplex the amine-aldehyde pair adopts a conformation which is unfavorable in solution, yet is completely isosteric to canonical base pairs and does not distort the DNA duplex. We therefore concluded that the need for this uncommon conformation might be the reason for the poor performance with the DNA polymerase. To verify our hypothesis, a naphthalene amine nucleoside was synthesized, which is almost a complete isostere of the observed conformation of the aromatic amine in the crystal structure, and was transformed to the triphosphate (dNaaTP). Indeed, this considerably improved the performance in primer extension experiments. Using the Terminator polymerase, faithful base pair formation and subsequent elongation was possible in the presence of dNaaTP and all canonical dNTPs. Based on this result, a further improved amine-aldehyde base pair was designed, which is intended to enhance enzyme acceptance by providing minor groove H-bond acceptors for the polymerase.

The S:T<sub>o</sub> or S:Naa pair represent the first candidates of a novel class of unnatural base pairs, which interact through reversible covalent bonding and are potential substrates for DNA polymerases. Further chemical optimization could result in an aldehyde-amine base pair, which is replicated in polymerase chain reactions with comparable fidelity and efficiency like the natural bases, in the near future. At the same time, this is the first example of a site-specific and reversible DNA crosslink. Complex DNA nanostructures for *in vivo* applications could be stabilized by incorporating the amine-aldehyde pairs.

Terminally phosphate modified nucleotides are important substrates in next-generation sequencing methods.<sup>[3, 103, 208]</sup> The Cu(I)-catalyzed alkyne-azide cycloaddition (CuAAC) has been used extensively to functionalize DNA postsynthetically.<sup>[87-88, 91, 93, 97]</sup> Despite its success, CuAACs had not been applied in the synthesis of phosphate modified nucleotides.

In the course of the project, a short and efficient synthesis of  $\gamma$ -labeled nucleotides was established, which provided access to a variety of phosphate modified nucleotides through Cu(I)-catalyzed click chemistry. When the labeled nucleotides were employed in primer extension experiments, we found that they are accepted by the Klenow fragment DNA polymerase. Also, a  $\gamma$ -alkyne and  $\gamma$ -fluorescein modified guanosine triphosphate was incorporated by the T7 RNA polymerase in *in vitro* transcriptions. This allowed site-specific labeling of the RNA transcript at the 5'-position. Inspired by the chemical strategy to introduce the modification, an  $\alpha$ -alkyne dGTP was prepared. Unfortunately, it was found to be chemically unstable and was therefore inappropriate for chemo-enzymatic labeling. Also efforts to provide postsynthetic 5'-oligonucleotide labeling with  $\gamma$ -alkyne ATP substrates by the T4 oligonucleotide kinase were unsuccessful. Instead of label-transfer, only phosphate group transfer was observed, like for the natural ATP, even when a hydrolytically more stable  $\gamma$ -alkyne ATP was employed. While the stability of the  $\gamma$ -modified phosphoramidates was sufficiently stable for RNA-labeling by transcription, the  $\alpha$ -alkyne dGTP hydrolyzed even under mild conditions. Especially for the application with hydrolytically active

---

enzymes, like e.g. kinases, it is advisable to have at least a phosphate ester or even phosphonate bond for connecting the label. Although the synthesis of these labeled compounds is much more complicated and time-consuming, the results presented in this thesis clearly indicate that such an approach is necessary to overcome the hydrolytic instability.

In conclusion, the CuAAC mediated synthesis of  $\gamma$ -labeled nucleotides is short, modular and the resulting compounds are potential candidates for next-generation sequencing approaches. Moreover,  $\gamma$ -labeled GTP allows site-specific labeling of RNA transcripts (5'-end) *in vitro*. Efforts to expand the labeling strategy to other systems, demonstrate the need for chemical redesign of the substrates and could also profit from engineered or evolved enzymes.

## 5 Experimental Part

### 5.1 General Methods and Materials for Synthesis

Chemicals were purchased from SIGMA ALDRICH, FLUKA, ABCR, ALFA AESAR or ACROS ORGANICS and used without further purification. Sensitive compounds like triphosphate derivatives were freeze-dried on a lyophilizer (CHRIST ALPHA 2-4 LD). All other solutions were concentrated *in vacuo* on a HEIDOLPH rotary evaporator with a Vario PC2001 diaphragm pump by VACUUBRAND. The solvents for organic synthesis were of reagent grade and purified by distillation. Dry solvents were bought from SIGMA-ALDRICH. Acetonitrile for HPLC purification was purchased from VWR. Water was purified by a Milli-Q Plus system from MERCK MILLIPORE.

All reactions were carried out with magnetic stirring, and if moisture and air sensitive, in oven-dried glassware (>12 h, 110 °C) under nitrogen. The temperature of reactions (except rt) was adjusted with a solvent/dry ice-, solvent/ice-mixture or an oil bath and the temperature monitored with a thermometer outside the flask.

Chromatographic purification of products was accomplished using flash column chromatography on MERCK Geduran Si 60 (40 – 63 µm) silica gel (normal phase) or by reversed-phase high-performance liquid chromatography (RP-HPLC). Thin layer chromatography (TLC) was performed on MERCK 60 (silica gel F<sub>254</sub>) plates and visualized under UV light ( $\lambda = 254$  and 366 nm) and staining with potassium permanganate (1.5 g KMnO<sub>4</sub>, 10.0 g K<sub>2</sub>CO<sub>3</sub>, 125 mg NaOH in 200 mL water) or ceric ammonium molybdate (10.0 g ammonium molybdate tetrahydrate, 2 g Ce(SO<sub>4</sub>)<sub>2</sub>·4H<sub>2</sub>O, 180 mL ddH<sub>2</sub>O, 20 mL conc. H<sub>2</sub>SO<sub>4</sub>).

<sup>1</sup>H, <sup>13</sup>C, <sup>19</sup>F and <sup>31</sup>P NMR spectra were recorded in deuterated solvents on VARIAN OXFORD 200, BRUKER ARX 300, VARIAN VXR400S, VARIAN INOVA 400, BRUKER Avance III (cryoprobe) 400, BRUKER AMX 600 and BRUKER Avance III HD (cryoprobe) 800 spectrometers and calibrated to the residual solvent peak using reported values.<sup>[222]</sup> As an external reference triphenyl phosphate (−18 ppm) was used for <sup>31</sup>P NMR spectra. The chemical shifts ( $\delta$ ) are given in ppm, the coupling constants ( $J$ ) in Hz. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad and combinations of these. For assignment of the structures, additional 2D NMR spectra (COSY, HSQC, HMBC) were measured. The numbering of atoms in a molecule in the experimental diagrams is depicted for clarity and does not correspond to IUPAC numbering.

Matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectra were recorded on a BRUKER AUTOFLEX II. For MALDI-TOF measurements the samples were dialyzed on a 0.025 µm VSWP filter (MILLIPORE) against ddH<sub>2</sub>O for 2 h and then co-crystallized in a 3-hydroxypicolinic acid matrix (HPA: 25 mg 3-hydroxypicolinic acid, 5 mg ammonium citrate, 5 µL 15-crown-5 in 0.5 mL H<sub>2</sub>O/MeCN = 1:1). High resolution electrospray ionization mass spectra (HRMS-ESI) were recorded on a THERMO FINNIGAN LTQ-FT (ESI-FTICR), and high resolution electron impact ionization mass spectra (HRMS-EI) were recorded on a THERMO FINNIGAN MAT 95. IR spectra were recorded on a PERKIN ELMER spectrum



BX instrument and are reported as follows: wavenumber  $\tilde{\nu}$  in  $\text{cm}^{-1}$ . The pH values of buffers were adjusted using a MP 220 pH meter (METTLER TOLEDO). UV spectra and melting profiles were measured on a JASCO V-650 spectrometer using quartz glass cuvettes with 1 cm path length. For fluorescence measurements a VARIAN Cary Eclipse spectrofluorometer was used. The extinction coefficient of the artificial nucleotides (dT<sub>o</sub>TP, dSTP, STP, dNaaTP) was determined by Beer-Lambert law from 4-5 adequate concentrations (absorbance 0.10-1.00) measured in triplicate on the UV-Vis spectrometer or on a NanoDrop ND-1000 Spectrophotometer (THERMO SCIENTIFIC).

### Reversed-Phase High-Performance Liquid Chromatography (RP-HPLC)

The following devices were used for the analysis and purification of the synthetic DNA/RNA strands, some nucleosides and the nucleoside (tri)phosphates. In order to remove insoluble particles, samples for preparative HPLC were filtered through a 0.2  $\mu\text{m}$  GHP filter membrane (Acrodisc<sup>®</sup>), and samples for analytical HPLC were centrifuged at maximum speed, prior to loading.

Analytical RP-HPLC was performed on an analytical HPLC WATERS Alliance (2695 Separation Module, 2996 Photodiode Array Detector) equipped with the column Nucleosil 120-3 C18 from MACHEREY NAGEL. Using a flow of 0.5 mL/min, gradients of 0–30% B in 45 min or 0–40% B in 45 min were applied for the oligonucleotides. Preparative RP-HPLC was performed on a HPLC WATERS Breeze (2487 Dual  $\lambda$  Array Detector, 1525 Binary HPLC Pump) equipped with the columns Nucleosil 100-7 C18, VP 250/10 C18 or VP 250/32 C18 from MACHEREY NAGEL. Using a flow of 5 mL/min, a gradient of 0–30% B was applied for 15 mer oligonucleotides and 0–40% B was applied for 30 mer oligonucleotides. For conditions of (tri)phosphate purification, see the individual chemical synthesis of the compounds.

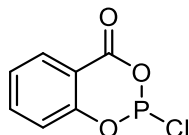
Buffer systems: Buffer A: 0.1 M triethylammonium acetate in water

Buffer B: 0.1 M triethylammonium acetate in 80% (v/v) acetonitrile

## 5.2 Chemical Synthesis

### 5.2.1 Preparation of the Triphosphate Reagent Solution

2-Chloro-4*H*-benzo[*d*][1,3,2]dioxaphosphinin-4-one (**10**)



$\text{C}_7\text{H}_4\text{ClO}_3\text{P}$   
MW: 202.53 g/mol

The synthesis is based on a modified procedure from *M. Donahue et al.*<sup>[190]</sup>

Salicylic acid was predried overnight by applying high vacuum ( $10^{-2}$  mbar), and then a part (3.45 g, 25.0 mmol, 1.0 eq) was transferred to a heat-dried round-bottom flask (50 mL) under an atmosphere of argon. Dry toluene (3.8 mL) and phosphorus trichloride (2.4 mL, 3.78 g, 27.5 mmol, 1.1 eq) were added and the mixture was refluxed at 110 °C for 2 h. The developing HCl gas was passed into a saturated aqueous solution of  $\text{NaHCO}_3$  for quenching. After cooling the pale green reaction solution to rt all volatiles were removed (and quenched) *in vacuo*. The resulting colorless solid (about 85% P(III) product **10** as determined by  $^{31}\text{P}$  NMR) was used without further purifications for the triphosphate reagent formation.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 8.03 (dd,  $J$  = 7.9, 1.7 Hz, 1H,  $\text{H}_{\text{ar}}$ ), 7.63 (ddd,  $J$  = 8.4, 7.5, 1.8 Hz, 1H,  $\text{H}_{\text{ar}}$ ), 7.27 (t,  $J$  = 7.7 Hz, 1H,  $\text{H}_{\text{ar}}$ ), 7.08 (d,  $J$  = 8.3 Hz, 1H,  $\text{H}_{\text{ar}}$ ).

$^{31}\text{P}$  NMR (162 MHz,  $^1\text{H}$ -decoupled,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 148.2.

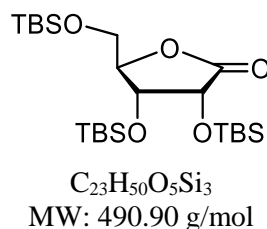
#### General Procedure 1: Triphosphate Reagent Solution

This procedure describes an exemplary setup for the generation of a triphosphate reagent solution, which can be applied to different nucleosides with a total amount of about 480  $\mu\text{mol}$  in parallel based on published protocols.<sup>[51-52, 64]</sup>

The phosphite 2-chloro-4*H*-benzo[*d*][1,3,2]dioxaphosphinin-4-one (**10**) and the tributylamine pyrophosphate  $((\text{TBA})_2\text{PP}_i)^{[50]}$  were dried under high vacuum for at least 3 h, the tributylamine was dried over molecular sieves 3 Å for at least 3 h prior to usage. All reagents were kept under an atmosphere of argon and the flasks (50 mL) were heat-dried. To a stirring solution of  $(\text{TBA})_2\text{PP}_i$  (550 mg, 1.0 mmol, 2.0 eq) in dry DMF (1.7 mL) was added tributylamine (1.86 mL, 1.45 g, 7.9 mmol, 15.8 eq) at rt. After 5 min, the resulting emulsion was poured into a solution of 2-chloro-4*H*-benzo[*d*][1,3,2]dioxaphosphinin-4-one (**10**, 203 mg, 1.0 mmol, 2.0 eq) in dry DMF (1.7 mL) and stirred for 30 min at rt. The resulting solution contains about 190 mM of a preformed triphosphate reagent, which is usually applied to the dry nucleoside at 0 °C using 2 equivalents of the reagent per equivalent of the nucleoside.

### 5.2.2 Synthesis of the Salicylaldehyde Ribonucleotide

(3*R*,4*R*,5*R*)-3,4-Bis(*tert*-butyldimethylsilyloxy)-5-((*tert*-butyldimethylsilyloxy)methyl)dihydrofuran-2(3*H*)-one (**67**)<sup>[167, 223]</sup>



$\beta$ -D-Ribose (6.0 g, 40 mmol, 1.0 eq) and  $\text{NaHCO}_3$  (6.7 g, 80 mmol, 2.0 eq) were dissolved in  $\text{dH}_2\text{O}$  (70 mL) and cooled to 0 °C. Bromine (2.2 mL, 43 mmol, 1.1 eq) was added dropwise to the stirring solution at 0 °C and the mixture was allowed to warm to rt slowly and stirred for 2 h. The excess bromine was quenched by titration with a solution of  $\text{Na}_2\text{S}_2\text{O}_3$  until complete color discharge, the solution was concentrated *in vacuo* (60 °C) and the residue was excessively dried for at least 20 h under high vacuum. The off-white solid containing ribolactone **85** was dissolved together with imidazol (23.4 g, 344 mmol, 8.6 eq) in dry DMF (120 mL) and cooled to 0 °C. Then, *tert*-butyldimethylsilyl chloride (21.2 g, 141 mmol, 3.5 eq) was added in one portion under stirring, the solution was allowed to warm to rt and finally heated to 50 °C for 3 h. The mixture was poured into  $\text{dH}_2\text{O}$  (250 mL) and was extracted with diethyl ether (3 x 150 mL). The combined organic layers were washed with a saturated aqueous solution of  $\text{NaHCO}_3$ , then  $\text{NaCl}$  and finally  $\text{dH}_2\text{O}$  (100 mL each), dried over  $\text{Na}_2\text{SO}_4$ , filtered and concentrated *in vacuo*. Flash column chromatography (silica, 18 x 10 cm, wet load, gradient; *i*Hex/EtOAc = 1:0  $\rightarrow$  100:1  $\rightarrow$  50:1) yielded the TBS-protected ribolactone **67** (11.7 g, 23.8 mmol, 60%) as a colorless solid.

$R_f$  (*i*Hex/EtOAc, 10:1) = 0.7.

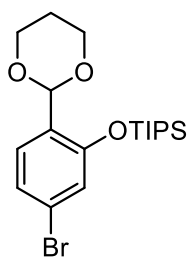
$^1\text{H NMR}$  (300 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 4.57 (d,  $J$  = 5.2 Hz, 1H), 4.26–4.23 (m, 2H), 3.84 (dd,  $J$  = 11.7, 3.0 Hz, 1H), 3.80 (dd,  $J$  = 11.7, 2.2 Hz, 1H), 0.94 (s, 9H), 0.89 (s, 9H), 0.88 (s, 9H), 0.18 (s, 3H), 0.13 (s, 3H), 0.10 (s, 3H), 0.09 (s, 3H), 0.07 (s, 3H), 0.06 (s, 3H).

$^{13}\text{C NMR}$  (75 MHz,  $^1\text{H}$ -decoupled,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 175.3, 85.8, 72.2, 70.8, 62.6, 26.1, 26.0, 25.9, 18.6, 18.5, 18.4, -4.4, -4.4, -4.6, -4.8, -5.3, -5.5.

**IR** (ATR):  $\tilde{\nu}$  ( $\text{cm}^{-1}$ ) = 2953, 2929, 2885, 2857, 2362, 2340, 1787, 1472, 1463, 1252, 1166, 1104, 984, 953, 832, 776.

**HRMS** (ESI): calc. for  $\text{C}_{23}\text{H}_{54}\text{O}_5\text{NSi}_3^+$  [ $\text{M}+\text{NH}_4$ ] $^+$ : 508.3304; found: 508.3302.

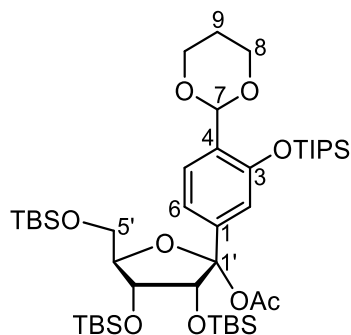
(5-Bromo-2-(1,3-dioxan-2-yl)phenoxy)triisopropylsilane (**2**)



$C_{19}H_{31}BrO_3Si$   
MW: 415.44 g/mol

Multigram synthesis (> 30 g of the salen base building block **2**) was performed according to the known and published procedures from *Clever et al.* with only minor adjustments.<sup>[12]</sup> NMR and mass data confirmed the identity of the compounds and were in agreement with the published data.

(2*R*,3*R*,4*R*,5*R*)-2-(4-(1,3-Dioxan-2-yl)-3-(triisopropylsilyloxy)phenyl)-3,4-bis(*tert*-butyldimethylsilyloxy)-5-((*tert*-butyldimethylsilyloxy)methyl)tetrahydrofuran-2-yl acetate (**69**)



$C_{44}H_{84}O_9Si_4$   
MW: 869.47 g/mol

Under an atmosphere of argon the protected salicyl base building block **2** (2.88 g, 6.93 mmol, 1.7 eq) was dissolved in dry diethyl ether (20 mL), degassed and cooled to  $-78\text{ }^{\circ}\text{C}$ . Then, *t*BuLi (1.6 M in pentane, 8.7 mL, 13.86 mmol, 3.4 eq) was slowly added dropwise within 25 min and the solution was stirred for 2 h at  $-78\text{ }^{\circ}\text{C}$ . TBS-protected ribolactone **67** (2.00 g, 4.08 mmol, 1.0 eq) dissolved in dry THF (16 mL) was added dropwise to the lithiated base. After 30 min, acetic anhydride (2.30 mL, 2.50 g, 24.48 mmol, 6.0 eq) was added and stirred for 15 min at  $-78\text{ }^{\circ}\text{C}$  before it was allowed to warm to rt. The mixture was poured into ice water (120 mL) and was extracted with diethyl ether (3 x 100 mL). The combined organic layers were dried over  $\text{Na}_2\text{SO}_4$ , filtered and concentrated *in vacuo* (at max.  $30\text{ }^{\circ}\text{C}$ ). Flash column chromatography (silica, 17 x 5.5 cm, wet load, gradient; *i*Hex/EtOAc = 1:0  $\rightarrow$  100:1  $\rightarrow$  50:1  $\rightarrow$  20:1) afforded the acetylated hemiketal **69** (2.45 g, 2.82 mmol, 69%) as a colorless solid.

$R_f$  (*i*Hex/EtOAc, 10:1) = 0.3.

$^1\text{H NMR}$  (599 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 7.50 (d,  $J$  = 8.1 Hz, 1H, 5-H), 7.12 (dd,  $J$  = 8.1 Hz,  $J$  = 1.7 Hz, 1H, 6-H), 6.82 (d,  $J$  = 1.7 Hz, 1H, 2-H), 5.83 (s, 1H, 7-H), 4.25 (ddd,  $J$  = 5.3, 3.2, 2.2 Hz, 1H, 4'-H), 4.24-

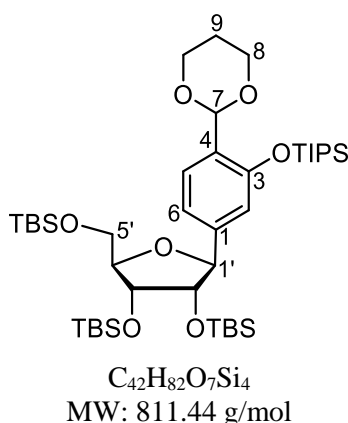
4.19 (m, 2H, 8-H), 4.15 (dd,  $J = 5.1, 2.1$  Hz, 1H, 3'-H), 3.97-3.89 (m, 2H, 8-H), 3.85 (dd,  $J = 10.8, 3.3$  Hz, 1H, 5'-H), 3.77-3.72 (m, 2H, 2'-H and 5'-H), 2.28-2.19 (m, 1H, 9-H), 2.02 (s, 3H, OCOCH<sub>3</sub>), 1.42-1.37 (m, 1H, 9-H), 1.32-1.25 (m, 3H, Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>), 1.14-1.09 (m, 18H, Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>), 0.93 (s, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.92 (s, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.90 (s, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.10 (s, 3H, SiCH<sub>3</sub>), 0.09 (s, 3H, SiCH<sub>3</sub>), 0.07-0.05 (m, 6H, 2 x SiCH<sub>3</sub>), -0.10 (s, 3H, SiCH<sub>3</sub>), -0.36 (s, 3H, SiCH<sub>3</sub>).

<sup>13</sup>C NMR (151 MHz, <sup>1</sup>H-decoupled, CDCl<sub>3</sub>):  $\delta$  (ppm) = 168.4 (OCOCH<sub>3</sub>), 152.7 (1/3/4-C)\*, 141.9 (1/3/4-C)\*, 128.6 (1/3/4-C)\*, 127.1 (5-C), 118.6 (6-C), 115.6 (2-C), 106.2 (1'-C), 97.5 (7-C), 86.7 (4'-C), 80.3 (2'-C), 71.9 (3'-C), 67.54 (8-C), 67.51 (8-C), 63.2 (5'-C), 26.0 (SiC(CH<sub>3</sub>)<sub>3</sub>), 25.9 (SiC(CH<sub>3</sub>)<sub>3</sub> and 9-C), 25.8 (SiC(CH<sub>3</sub>)<sub>3</sub>), 22.1 (OCOCH<sub>3</sub>), 18.1 (Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>), 13.0 (Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>), -4.2 (SiCH<sub>3</sub>), -4.5 (SiCH<sub>3</sub>), -4.6 (SiCH<sub>3</sub>), -5.3 (SiCH<sub>3</sub>), -5.4 (SiCH<sub>3</sub>), -5.5 (SiCH<sub>3</sub>).

IR (ATR):  $\tilde{\nu}$  (cm<sup>-1</sup>) = 2950, 2929, 2858, 2894, 1758, 1472, 1463, 1416, 1390, 1363, 1251, 1236, 1101, 1000, 950, 940, 880, 834, 774.

HRMS (ESI): calc. for C<sub>42</sub>H<sub>81</sub>O<sub>7</sub>Si<sub>4</sub><sup>+</sup> [oxocarbenium from M]<sup>+</sup>: 809.5054; found: 809.5058.

((2*S*,3*S*,4*R*,5*R*)-2-(4-(1,3-Dioxan-2-yl)-3-(triisopropylsilyloxy)phenyl)-5-((*tert*-butyldimethylsilyloxy)methyl)tetrahydrofuran-3,4-diyl)bis(oxy)bis(*tert*-butyldimethylsilane) (**66**)



To a stirring solution of the acetylated hemiketal **69** (2.20 g, 2.53 mmol, 1.0 eq) in dry toluene (16.5 mL) was added triethylsilane (1.21 mL, 0.88 g, 7.60 mmol, 3.0 eq) at -30 °C. After 5 min stirring the solution was warmed to -20 °C, BF<sub>3</sub>·OEt<sub>2</sub> (0.16 mL, 1.26 mmol, 0.5 eq) was added dropwise and stirring continued for 5 min. The deep dark green solution was quenched with an aqueous saturated solution of NaHCO<sub>3</sub> (12.5 mL) and diluted with dH<sub>2</sub>O (12.5 mL) and warmed to rt. The mixture was immediately extracted with diethyl ether (3 x 100 mL) and the combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and dried *in vacuo*. Flash column chromatography (silica, 20 x 5.5 cm, wet load, gradient; *i*Hex/EtOAc = 1:0 → 100:1 → 50:1 → 25:1) afforded the fully protected salicyl nucleoside **66** (0.63 g, 0.78 mmol, 30%) as a colorless wax.

**R<sub>f</sub>** (*i*Hex/EtOAc, 10:1) = 0.7.

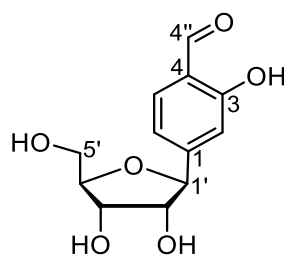
**<sup>1</sup>H NMR** (599 MHz, CDCl<sub>3</sub>):  $\delta$  (ppm) = 7.51 (d,  $J$  = 7.9 Hz, 1H, 5-H), 7.04 (dd,  $J$  = 8.0 Hz,  $J$  = 1.4 Hz, 1H, 6-H), 6.77 (d,  $J$  = 1.2 Hz, 1H, 2-H), 5.85 (s, 1H, 7-H), 4.74 (d,  $J$  = 5.9 Hz, 1H, 1'-H), 4.28-4.18 (m, 2H, 8-H), 4.07 (t,  $J$  = 3.9, 1H, 3'-H), 4.05-4.01 (m, 1H, 4'-H), 3.97-3.91 (m, 2H, 8-H), 3.80-3.76 (m, 2H, 2'-H and 5'-H), 3.72 (dd,  $J$  = 10.7, 5.5 Hz, 1H, 5'-H), 2.30-2.18 (m, 1H, 9-H), 1.45-1.37 (m, 1H, 9-H), 1.31 (hept,  $J$  = 7.7 Hz, 3H, Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>), 1.12 (dd,  $J$  = 7.5, 2.7 Hz, 18H, Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>), 0.94 (s, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.92 (s, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.85 (s, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.11 (s, 3H, SiCH<sub>3</sub>), 0.10 (s, 3H, SiCH<sub>3</sub>), 0.07 (s, 3H, SiCH<sub>3</sub>), 0.05 (s, 3H, SiCH<sub>3</sub>), -0.09 (s, 3H, SiCH<sub>3</sub>), -0.24 (s, 3H, SiCH<sub>3</sub>).

**<sup>13</sup>C NMR** (151 MHz, <sup>1</sup>H-decoupled, CDCl<sub>3</sub>):  $\delta$  (ppm) = 152.9 (1/3/4-C)\*, 142.7 (1/3/4-C)\*, 128.1 (1/3/4-C)\*, 126.9 (5-C), 119.6 (6-C), 115.8 (2-C), 97.5 (7-C), 84.2 (4'-C), 83.6 (1'-C), 79.3 (2'-C), 73.1 (3'-C), 67.6 (8-C), 63.9 (5'-C), 26.1 (SiC(CH<sub>3</sub>)<sub>3</sub>), 25.9 (2 x SiC(CH<sub>3</sub>)<sub>3</sub>), 25.8 (9-C), 18.1 (Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>), 13.1 (Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>), -4.4 (SiCH<sub>3</sub>), -4.6 (SiCH<sub>3</sub>), -4.6 (SiCH<sub>3</sub>), -4.8 (SiCH<sub>3</sub>), -5.4 (2 x SiCH<sub>3</sub>).

**IR** (ATR):  $\tilde{\nu}$  (cm<sup>-1</sup>) = 2950, 2929, 2894, 2858, 1472, 1464, 1427, 1390, 1361, 1252, 1150, 1123, 1100, 1081, 1003, 988, 939, 881, 832, 814, 774, 680, 671.

**HRMS (ESI)**: calc. for C<sub>42</sub>H<sub>83</sub>O<sub>7</sub>Si<sub>4</sub><sup>+</sup> [M+H]<sup>+</sup>: 811.5210; found: 811.5219.

4-((2*S*,3*R*,4*S*,5*R*)-3,4-Dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2-hydroxybenzaldehyde (**65**)



C<sub>12</sub>H<sub>14</sub>O<sub>6</sub>  
MW: 254.24 g/mol

**Route A:** To a stirring solution of the fully protected salicyl nucleoside **66** (600 mg, 0.74 mmol, 1.0 eq) in THF (9.1 mL) in a 50 mL polypropylene tube, was added HF·pyridine (0.77 mL, 29.6 mmol, 40.0 eq) at 0 °C. The solution was allowed to warm to rt and stirred for 1.5 h before dH<sub>2</sub>O (78  $\mu$ L, 4.3 mmol, 5.8 eq) was added and stirring was continued for 24 h. Me<sub>3</sub>SiOMe (8.4 mL, 61.1 mmol, 82.5 eq) was added and stirring was continued for 30 min. The solvent was removed *in vacuo* and flash column chromatography (silica, 16 x 4 cm, wet load, gradient; DCM/MeOH = 1:0  $\rightarrow$  50:1  $\rightarrow$  9:1) afforded the salicylaldehyde nucleoside **65** (83 mg, 0.33 mmol, 44%) as a yellow solid after excessive drying under high vacuum.

**Route B:** To a stirring solution of the acetylated hemiketal **69** (2.49 g, 2.86 mmol, 1.0 eq) in dry toluene (18.8 mL) at -30 °C was added triethylsilane (1.37 mL, 8.58 mmol, 3.0 eq). After 5 min stirring the solution was warmed to -20 °C, BF<sub>3</sub>·OEt<sub>2</sub> (0.18 mL, 1.43 mmol, 0.5 eq) was added dropwise and stirring was continued for 5 min. The deep dark green solution was quenched with an aqueous saturated solution of NaHCO<sub>3</sub> (13.5 mL) and diluted with ddH<sub>2</sub>O (13.5 mL) and warmed to rt. The mixture was immediately

extracted with diethyl ether (3 x 70 mL) and the combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and dried *in vacuo*. The crude product mixture, which contained the fully protected nucleoside **66** (about 50% determined by <sup>1</sup>H NMR) was dissolved in THF (17.6 mL) transferred to a 50 mL polypropylene tube and cooled to 0 °C. HF·pyridine (1.04 mL, 114.4 mmol, 40.0 eq) was added at 0 °C and the solution was warmed to rt. After 3 h at rt, dH<sub>2</sub>O (0.15 mL, 16.6 mmol, 5.8 eq) was added and stirring continued for 72 h. The solution was quenched with Me<sub>3</sub>SiOMe (16.3 mL, 236.6 mmol, 82.5 eq) at rt within 30 min and concentrated *in vacuo*. Flash column chromatography (silica, 16 x 4 cm, wet load, gradient; DCM/MeOH = 1:0 → 9:1) afforded the salicylaldehyde nucleoside **65** as a mixture with 1,3-propanediol (370 mg, about 205 mg ribonucleoside **65**, 28% from hemiketal **69** based on <sup>1</sup>H NMR) as a yellow oil. Preparative RP-HPLC purification (0–30% acetonitrile in ddH<sub>2</sub>O in 45 min) afforded the salicylaldehyde nucleoside **65** (145 mg, 0.57 mmol, 20% from acetylated hemiketal **69**) as a dark yellow solid.

**R<sub>f</sub>** (DCM/MeOH, 9:1) = 0.2.

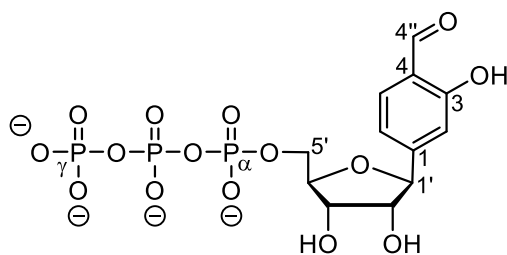
**<sup>1</sup>H NMR** (400 MHz, CD<sub>3</sub>OD): δ (ppm) = 9.98 (s, 1H, 4''-H), 7.64 (d, <sup>3</sup>J = 8.0 Hz, 1H, 5-H), 7.10 (ddd, J = 8.0, 1.5, 0.5 Hz, 1H, 6-H), 7.07–7.05 (m, 1H, 2-H), 4.69 (d, <sup>3</sup>J = 6.9 Hz, 1H, 1'-H), 4.04–4.00 (m, 1H, 3'-H), 4.00–3.95 (m, 1H, 4'-H), 3.84–3.80 (m, 1H, 2'-H), 3.80–3.69 (m, 2H, 5'-H).

**<sup>13</sup>C NMR** (101 MHz, <sup>1</sup>H-decoupled, CD<sub>3</sub>OD): δ (ppm) = 196.7 (4''-C), 163.2 (3-C), 152.2 (1-C), 133.7 (5-C), 122.1 (4-C), 118.3 (6-C), 115.6 (2-C), 86.6 (4'-C), 84.7 (1'-C), 79.2 (2'-C), 73.0 (3'-C), 63.5 (5'-C).

**IR** (ATR):  $\tilde{\nu}$  (cm<sup>-1</sup>) = 3445, 3312, 3211, 2972, 2950, 2936, 2876, 1658, 1626, 1566, 1498, 1436, 1382, 1345, 1304, 1223, 1210, 1189, 1159, 1117, 1106, 1074, 1049, 1028, 1007, 979, 951, 902, 872, 845, 806, 782, 747, 732, 723, 694, 677, 639.

**HRMS (ESI)**: calc. for C<sub>12</sub>H<sub>13</sub>O<sub>6</sub><sup>-</sup> [M-H]<sup>-</sup>: 253.0718; found: 253.0721.

((2*R*,3*S*,4*R*,5*S*)-5-(4-Formyl-3-hydroxyphenyl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl triphosphate (**1**)



C<sub>12</sub>H<sub>17</sub>O<sub>15</sub>P<sub>3</sub> (hypothetic free acid)  
MW: 494.18 g/mol

Triphosphate reagent solution (0.84 mL, 160 μmol, 2.0 eq, prepared according to general procedure **1**, p. 100) was added to the excessively dried salicylaldehyde nucleoside **65** (20.3 mg, 80 μmol, 1.0 eq) in an argon atmosphere at rt. After 90 min an iodine solution (20 mM I<sub>2</sub> in Py/H<sub>2</sub>O = 9:1) was added until a slight yellow-brown color remained for 15 min (approx. 0.7 mL), followed by the addition of ddH<sub>2</sub>O (2.0 mL), and stirring was continued at rt for 1.5 h. An aqueous solution of NaCl (3 M, 0.5 mL) was added, and the solution was transferred to a 50 mL centrifugation tube, vortexed vigorously and mixed with EtOH

(absolute, 15 mL). Precipitation was achieved by cooling to  $-80\text{ }^{\circ}\text{C}$  for 45 min, followed by centrifugation (10 min,  $3200 \times g$ ). The supernatant was discarded and the pellet was redissolved in buffer A (1.0 mL) and lyophilized. The residue was purified twice by RP-HPLC (0–20% B in 65 min). This afforded the salicylaldehyde triphosphate **1** tetrakis-triethylammonium salt as a yellow solid (10.9 mg,  $10.4\text{ }\mu\text{mol}$ , 13%) after freeze-drying.

**$^1\text{H}$  NMR** (400 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = 9.86 (s, 1H, 4''-H), 7.65 (d,  $J=7.9\text{ Hz}$ , 1H, 5-H), 7.09–7.01 (m, 2H, 6-H and 2-H), 4.69–4.66 (m, 1H, 1'-H), 4.27–4.21 (m, 1H, 3'-H), 4.18–4.12 (s, 1H, 4'-H), 4.12–4.07 (m, 2H, 5'-H), 4.05–3.97 (m, 1H, 2'-H).

**$^{31}\text{P}$  NMR** (162 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) =  $-11.07$  (d,  $^2J_{\text{PP}} = 19.8\text{ Hz}$ ,  $\gamma\text{-P}$ ),  $-11.41$  (dt,  $^2J_{\text{PP}} = 19.7\text{ Hz}$ ,  $^3J_{\text{PH}} = 4.3\text{ Hz}$ ,  $\alpha\text{-P}$ ),  $-23.54$  (t,  $^2J_{\text{PP}} = 20.0\text{ Hz}$ ,  $\beta\text{-P}$ ).

**HRMS (ESI)**: calc. for  $\text{C}_{12}\text{H}_{16}\text{O}_{15}\text{P}_3^-$  [ $\text{M-H}$ ] $^-$ : 492.97075; found: 492.97086.

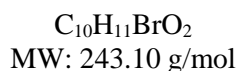
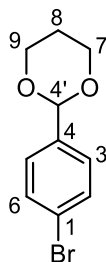
**UV-Vis** ( $\text{H}_2\text{O}$ ):  $\lambda_{\text{Abs}}$  (nm) = 289 (base peak), 233, 330.

**Extinction coefficient** ( $\text{H}_2\text{O}$ ):  $\varepsilon$  (289 nm) =  $13500\text{ M}^{-1}\text{cm}^{-1}$ ,  $\varepsilon$  (260 nm) =  $10300\text{ M}^{-1}\text{cm}^{-1}$ .



### 5.2.3 Synthesis of the Benzaldehyde Ribonucleotide

#### 2-(4-Bromophenyl)-1,3-dioxane (**71**)



To a stirring solution of 4-bromobenzaldehyde (**72**, 7.00 g, 35.28 mmol, 1.0 eq) in toluene (44 mL) 1,3-propanediol (3.00 mL, 41.92 mmol, 1.2 eq) and  $\text{BF}_3 \cdot \text{OEt}_2$  (0.11 mL, 0.86 mmol, 3-mol%) were added and the solution was heated to reflux (111 °C) for 4.5 h. After cooling to rt, half-saturated aqueous  $\text{NaHCO}_3$  (140 mL) was added and the organic layer was separated. The aqueous layer was extracted with toluene (2 x 75 mL) and the combined organic layers were dried over  $\text{MgSO}_4$  and concentrated *in vacuo*. Dioxane **71** was yielded as a colorless solid (8.35 g, 34.3 mmol, 97%) without further purifications.

$R_f$  (*i*Hex/EtOAc, 10:1) = 0.4.

**$^1\text{H}$  NMR** (599 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 7.52–7.46 (m, 2H, 2-H and 6-H), 7.38–7.33 (m, 2H, 3-H and 5-H), 5.46 (s, 1H, 4'-H), 4.29–4.23 (m, 2H, 7a-H and 9a-H), 4.01–3.94 (m, 2H, 7b-H and 9b-H), 2.26–2.18 (m, 1H, 8a-H), 1.45 (ddt,  $J$  = 13.5, 2.6, 1.2 Hz, 1H, 8b-H).

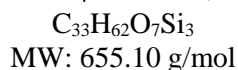
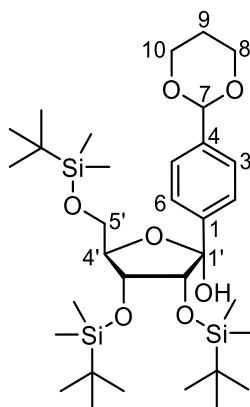
**$^{13}\text{C}$  NMR** (151 MHz,  $^1\text{H}$ -decoupled,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 137.9 (1-C), 131.50 (2-C and 6-C), 127.94 (3-C and 5-C), 122.96 (4-C), 100.96 (4'-C), 67.51 (7-C and 9-C), 25.83 (8-C).

**mp.**: 59 °C.

**IR** (ATR):  $\tilde{\nu}$  ( $\text{cm}^{-1}$ ) = 2969, 1592, 1382, 1100, 1010, 800.

**MS (ESI)**: calc. for  $\text{C}_{10}\text{H}_{12}^{79}\text{BrO}_2^+ [\text{M}+\text{H}]^+$ : 243.00; found: 242.93.

(2*S*,3*R*,4*R*,5*R*)-2-(4-(1,3-Dioxan-2-yl)phenyl)-3,4-bis(*tert*-butyldimethylsilyloxy)-5-((*tert*-butyldimethylsilyloxy)methyl)tetrahydrofuran-2-ol (**73**)



To a degassed stirring solution of the bromo-dioxane **71** (2.68 g, 11.0 mmol, 1.7 eq) in dry THF (31 mL) at  $-78\text{ }^{\circ}\text{C}$  under an atmosphere of argon was added *t*BuLi (1.5 M in pentane, 13.8 mL, 22.1 mmol, 3.4 eq.) dropwise within 30 min. After 2 h stirring at  $-78\text{ }^{\circ}\text{C}$ , a solution of the TBS-protected ribolactone **67** (3.18 g, 6.5 mmol, 1.0 eq) in dry, degassed THF (26.0 mL) was added dropwise and stirring was continued for 1 h. A saturated aqueous solution of  $\text{NH}_4\text{Cl}$  (4 mL) was added and the mixture was allowed to warm to rt. The mixture was diluted with  $\text{dH}_2\text{O}$  (100 mL) and extracted with diethyl ether (3 x 160 mL). The combined organic layers were dried over  $\text{Na}_2\text{SO}_4$  and concentrated *in vacuo*. Flash column chromatography (silica, 25 x 6.0 cm, wet load, gradient; *i*Hex/EtOAc = 1:0  $\rightarrow$  100:1  $\rightarrow$  50:1  $\rightarrow$  20:1  $\rightarrow$  10:1  $\rightarrow$  5:1) afforded the desired hemiketal **73** (3.88 g, 6.1 mmol, 91%) as a colorless solid.

$R_f$  (*i*Hex/EtOAc, 10:1) = 0.3.

$^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 7.63 (d,  $J$  = 8.3 Hz, 2H, 2-H and 6-H), 7.43-7.38 (d,  $J$  = 8.4 Hz, 2H, 3-H and 5-H), 5.49 (s, 1H, 7-H), 5.05 (s, 1H, 1'-OH), 4.25 (ddd,  $J$  = 11.9, 5.0, 1.5 Hz, 2H, 8a-H and 10a-H), 4.20 (t,  $J$  = 3.2 Hz, 1H, 4'-H), 4.18 (dd,  $J$  = 4.6, 1.0 Hz, 1H, 3'-H), 4.01 (d,  $J$  = 4.7 Hz, 1H, 2'-H), 4.00-3.93 (m, 2H, 8b-H and 10b-H), 3.88-3.75 (m, 2H, 5'-H), 2.29-2.15 (m, 1H, 9a-H), 1.46-1.39 (m, 1H, 9b-H), 0.93 (s, 9H,  $\text{SiC}(\text{CH}_3)_3$ ), 0.92 (s, 9H,  $\text{SiC}(\text{CH}_3)_3$ ), 0.81 (s, 9H,  $\text{SiC}(\text{CH}_3)_3$ ), 0.12 (s, 3H,  $\text{SiCH}_3$ ), 0.11 (s, 6H, 2 x  $\text{SiCH}_3$ ), 0.10 (s, 3H,  $\text{SiCH}_3$ ),  $-0.19$  (s, 3H,  $\text{SiCH}_3$ ),  $-0.55$  (s, 3H,  $\text{SiCH}_3$ ).

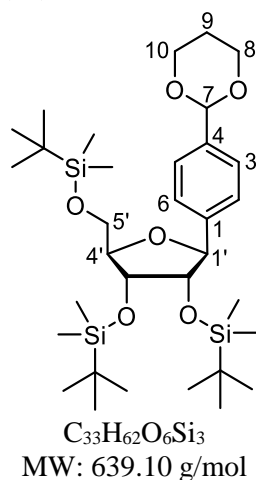
$^{13}\text{C}$  NMR (101 MHz,  $^1\text{H}$ -decoupled,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 141.6 (1-C), 138.6 (4-C), 126.9 (2-C and 6-C), 125.3 (3-C and 5-C), 104.1 (1'-C), 101.8 (7-C), 85.0 (4'-C), 78.1 (2'-C), 74.9 (3'-C), 67.49 (8-C and 10-C), 63.7 (5'-C), 25.92 (2 x  $\text{SiC}(\text{CH}_3)_3$ ), 25.89 ( $\text{SiC}(\text{CH}_3)_3$ ), 18.45 ( $\text{SiC}(\text{CH}_3)_3$ ), 18.03 ( $\text{SiC}(\text{CH}_3)_3$ ), 17.97 ( $\text{SiC}(\text{CH}_3)_3$ ),  $-4.44$  ( $\text{SiCH}_3$ ),  $-4.49$  ( $\text{SiCH}_3$ ),  $-4.68$  ( $\text{SiCH}_3$ ),  $-5.29$  ( $\text{SiCH}_3$ ),  $-5.5$  ( $\text{SiCH}_3$ ),  $-5.6$  ( $\text{SiCH}_3$ ).

mp.:  $69\text{ }^{\circ}\text{C}$ .

IR (ATR):  $\tilde{\nu}$  ( $\text{cm}^{-1}$ ) = 2929, 1472, 1252, 1108, 833.

HRMS (ESI): calc. for  $\text{C}_{33}\text{H}_{61}\text{O}_7\text{Si}_3^-$  [ $\text{M}-\text{H}$ ] $^-$ : 653.3731; found: 653.3733.

((2*S*,3*S*,4*R*,5*R*)-2-(4-(1,3-Dioxan-2-yl)phenyl)-5-((*tert*-butyldimethylsilyloxy)methyl)tetrahydrofuran-3,4-diyl)bis(oxy)bis(*tert*-butyldimethylsilane) (**74**)



Triethylsilane (2.32 mL, 14.5 mmol, 3.3 eq.) was added to a stirring solution of hemiketal **73** (2.88 g, 4.40 mmol, 1.0 eq.) in dry DCM (15.9 mL) at  $-78\text{ }^{\circ}\text{C}$ . After 10 min,  $\text{BF}_3\cdot\text{OEt}_2$  (0.66 mL, 5.28 mmol,

1.2 eq.) was added dropwise and stirring was continued for 1 h. A half-saturated aqueous solution of  $\text{NaHCO}_3$  (43 mL) was added, the mixture was warmed to rt and extracted with DCM (3 x 75 mL). The combined organic layers were dried over  $\text{Na}_2\text{SO}_4$ , filtered and concentrated *in vacuo*. Flash column chromatography (silica, 25 x 6.0 cm, wet load, gradient; *i*Hex/EtOAc = 1:0  $\rightarrow$  100:1  $\rightarrow$  50:1  $\rightarrow$  20:1  $\rightarrow$  10:1  $\rightarrow$  5:1) afforded the desired TBS-protected acetal nucleoside **74** (2.61 g, 4.1 mmol, 93%) as a colorless solid and also TBS-protected benzaldehyde nucleoside **77** (0.12 g, 0.2 mmol, 4%).

$R_f$  (*i*Hex/EtOAc, 10:1) = 0.5.

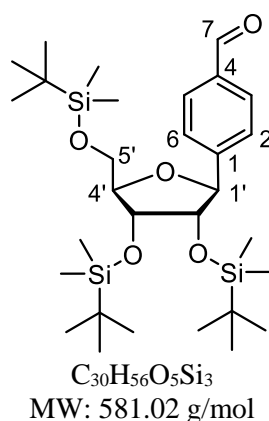
**$^1\text{H}$  NMR** (599 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 7.44-7.40 (m, 4H, 2-, 3-, 5- and 6-H), 5.49 (s, 1H, 7-H), 4.79 (d,  $J$  = 6.8 Hz, 1H, 1'-H), 4.27 (ddd,  $J$  = 11.9, 5.0, 1.5 Hz, 2H, 8a-H and 10a-H), 4.14-4.12 (m, 1H, 3'-H), 4.03-4.01 (m, 1H, 4'-H), 4.01-3.95 (m, 2H, 8b-H and 10b-H), 3.86-3.82 (m, 1H, 2'-H), 3.83-3.76 (m, 2H, 5'-H), 2.29-2.18 (m, 1H, 9a-H), 1.44 (ddt,  $J$  = 13.5, 2.6, 1.2 Hz, 1H, 9b-H), 0.94 (s, 9H,  $\text{SiC}(\text{CH}_3)_3$ ), 0.92 (s, 9H,  $\text{SiC}(\text{CH}_3)_3$ ), 0.82 (s, 9H,  $\text{SiC}(\text{CH}_3)_3$ ), 0.12 (s, 3H,  $\text{SiCH}_3$ ), 0.11 (s, 3H,  $\text{SiCH}_3$ ), 0.08 (s, 3H,  $\text{SiCH}_3$ ), 0.07 (s, 3H,  $\text{SiCH}_3$ ), -0.13 (s, 3H,  $\text{SiCH}_3$ ), -0.38 (s, 3H,  $\text{SiCH}_3$ ).

**$^{13}\text{C}$  NMR** (151 MHz,  $^1\text{H}$ -decoupled,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 141.6 (1-C), 138.2 (4-C), 126.8 (2-C and 6-C), 125.8 (3-C and 5-C), 101.8 (7-C), 85.4 (4'-C), 83.4 (1'-C), 79.7 (2'-C), 73.5 (3'-C), 67.5 (8-C and 10-C), 63.6 (5'-C), 26.2 ( $\text{SiC}(\text{CH}_3)_3$ ), 26.1 (2 x  $\text{SiC}(\text{CH}_3)_3$ ), 25.9 (9-C), 18.6 ( $\text{SiC}(\text{CH}_3)_3$ ), 18.2 ( $\text{SiC}(\text{CH}_3)_3$ ), 18.1 ( $\text{SiC}(\text{CH}_3)_3$ ), -4.25 ( $\text{SiCH}_3$ ), -4.32 ( $\text{SiCH}_3$ ), -4.4 ( $\text{SiCH}_3$ ), -4.9 ( $\text{SiCH}_3$ ), -5.2 ( $\text{SiCH}_3$ ), -5.3 ( $\text{SiCH}_3$ ).

**IR** (ATR):  $\tilde{\nu}$  ( $\text{cm}^{-1}$ ) = 2858, 1593, 1471, 1382, 1102, 1010, 834.

**HRMS (ESI)**: calc. for  $\text{C}_{33}\text{H}_{63}\text{O}_6\text{Si}_3^+$   $[\text{M}+\text{H}]^+$ : 639.3927; found: 639.3921.

4-((2*S*,3*S*,4*R*,5*R*)-3,4-Bis(*tert*-butyldimethylsilyloxy)-5-((*tert*-butyldimethylsilyloxy)methyl)tetrahydrofuran-2-yl)benzaldehyde (**77**)



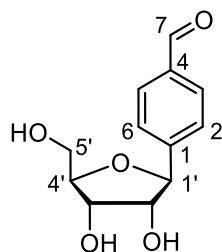
To a stirring solution of the acetal TBS nucleoside **74** (0.66 g, 1.04 mmol, 1.0 eq) in THF (6.3 mL) was added aqueous HCl (2 M, 0.69 mL) at rt. After 1 h, the reaction solution was neutralized with aqueous  $\text{NaHCO}_3$ , diluted with  $\text{dH}_2\text{O}$  (50 mL) and extracted with diethyl ether (3 x 100 mL). The combined organic layers were dried over  $\text{MgSO}_4$ , filtered and concentrated *in vacuo*. Flash column chromatography (silica, 20 x 3.0 cm, wet load, gradient; *i*Hex/EtOAc = 1:0  $\rightarrow$  100:1  $\rightarrow$  50:1  $\rightarrow$  20:1  $\rightarrow$  10:1  $\rightarrow$  5:1) afforded the TBS-protected benzaldehyde nucleoside **77** (0.52 g, 0.90 mmol, 86%) as a colorless solid.

$R_f$  (iHex/EtOAc, 10:1) = 0.6.

$^1\text{H NMR}$  (599 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 10.00 (s, 1H, 7-H), 7.83 (d,  $J$  = 8.0 Hz, 2H, 2-H and 6-H), 7.61 (d,  $J$  = 8.4 Hz, 2H, 3-H and 5-H), 4.84 (d,  $J$  = 8.0 Hz, 1H, 1'-H), 4.16-4.11 (m, 1H, 3'-H), 4.07-4.05 (m, 1H, 4'-H), 3.87 (dd,  $J$  = 8.0, 4.4 Hz, 1H, 2'-H), 3.83 (dd,  $J$  = 11.0, 3.8 Hz, 1H, 5'-H), 3.79 (dd,  $J$  = 11.0, 2.7 Hz, 1H, 5'-H), 0.94 (s, 18H, 2 x  $\text{SiC}(\text{CH}_3)_3$ ), 0.81 (s, 9H,  $\text{SiC}(\text{CH}_3)_3$ ), 0.14 (s, 3H,  $\text{SiCH}_3$ ), 0.12 (s, 3H,  $\text{SiCH}_3$ ), 0.10 (s, 3H,  $\text{SiCH}_3$ ), 0.10 (s, 3H,  $\text{SiCH}_3$ ), -0.12 (s, 3H,  $\text{SiCH}_3$ ), -0.50 (s, 3H,  $\text{SiCH}_3$ ).

$^{13}\text{C NMR}$  (151 MHz,  $^1\text{H}$ -decoupled,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 192.3 (7-C), 148.3 (4-C), 136.1 (1-C), 129.7 (2-C and 6-C), 127.6 (3-C and 5-C), 86.7 (4'-C), 82.5 (1'-C), 79.9 (2'-C), 74.3 (3'-C), 63.8 (5'-C), 26.1 ( $\text{SiC}(\text{CH}_3)_3$ ), 26.04 ( $\text{SiC}(\text{CH}_3)_3$ ), 25.98 ( $\text{SiC}(\text{CH}_3)_3$ ), 18.5 ( $\text{SiC}(\text{CH}_3)_3$ ), 18.2 ( $\text{SiC}(\text{CH}_3)_3$ ), 18.1 ( $\text{SiC}(\text{CH}_3)_3$ ), -4.2 ( $\text{SiCH}_3$ ), -4.29 ( $\text{SiCH}_3$ ), -4.32 ( $\text{SiCH}_3$ ), -5.26 ( $\text{SiCH}_3$ ), -5.34 (2 x  $\text{SiCH}_3$ ).

4-((2*S*,3*R*,4*S*,5*R*)-3,4-Dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)benzaldehyde (**76**)



$\text{C}_{12}\text{H}_{14}\text{O}_5$   
MW: 238.24 g/mol

To a stirring solution of TBS-protected benzaldehyde nucleoside **77** (0.30 g, 0.52 mmol, 1.0 eq) in THF (6.4 mL) in a 50 mL polypropylene tube was added HF·pyridine (284  $\mu\text{L}$ , 15.60 mmol, 30.0 eq.) at 0 °C. The mixture was allowed to warm to rt and after 24 h,  $\text{dH}_2\text{O}$  (54  $\mu\text{L}$ , 3.02 mmol, 5.8 eq) was added. Stirring was continued for 48 h,  $\text{Me}_3\text{SiOMe}$  (4.48 mL, 32.5 mmol, 62.5 eq.) was added and after 30 min the mixture was concentrated *in vacuo*. Flash column chromatography (silica, 20 x 2.0 cm, wet load, gradient; DCM/MeOH = 1:0  $\rightarrow$  100:1  $\rightarrow$  50:1  $\rightarrow$  20:1  $\rightarrow$  10:1  $\rightarrow$  5:1) and subsequent RP-HPLC (0–30% B in 45 min) afforded the benzaldehyde nucleoside **76** (40 mg, 0.17 mmol, 33%) as a colorless solid. In addition, single and double TBS-protected benzaldehyde nucleosides were isolated in 52% combined yield.

$R_f$  (DCM/MeOH, 9:1) = 0.3.

$^1\text{H NMR}$  (400 MHz,  $\text{CD}_3\text{OD}$ ):  $\delta$  (ppm) = 9.97 (s, 7-H, about 70%), 7.89 (dt,  $J$  = 8.0, 0.3 Hz, 2H, 3-H and 5-H), 7.71-7.66 (m, 2H, 2-H and 6-H), 7.45 (s, 7-H, hydrated aldehyde form, about 30%), 4.79 (d,  $^3J$  = 7.0 Hz, 1H, 1'-H), 4.08-3.99 (m, 2H, 3'-H and 4'-H), 3.88-3.82 (m, 1H, 2'-H), 3.84-3.71 (m, 2H, 5'-H).

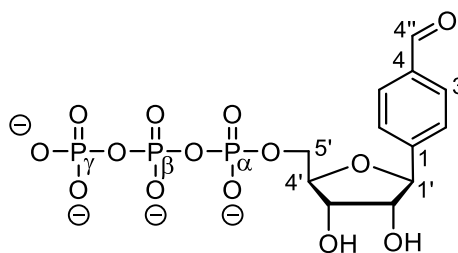
$^{13}\text{C NMR}$  (101 MHz,  $^1\text{H}$ -decoupled,  $\text{CD}_3\text{OD}$ ):  $\delta$  (ppm) = 193.9 (7-C), 149.5 (4-C), 137.4 (1-C), 130.7 (3-C and 5-C), 127.8 (2-C and 6-C), 127.0 (hydrated form 7-C), 86.8 (4'-H), 84.7 (1'-H), 79.3 (2'-H), 73.0 (3'-H), 63.5 (5'-H).

mp.: 138.5 °C.

IR (ATR):  $\tilde{\nu}$  ( $\text{cm}^{-1}$ ) = 3313, 2848, 2456, 1682, 1606, 1210, 1042, 815.

**HRMS (ESI):** calc. for  $C_{12}H_{13}O_5^- [M-H]^-$ : 237.0768; found: 237.0770.

((2*R*,3*S*,4*R*,5*S*)-5-(4-Formylphenyl)-3,4-dihydroxytetrahydrofuran-2-yl)methyl triphosphate (**78**)



$C_{12}H_{17}O_{14}P_3$  (hypothetic free acid)  
MW: 478.18 g/mol

The excessively dried benzaldehyde nucleoside (40.8 mg, 170  $\mu$ mol, 1.0 eq) was dissolved in dry DMF (0.82 mL) and the triphosphate reagent solution (1.79 mL, 340  $\mu$ mol, 2.0 eq, prepared according to general procedure **1**, p. 100) was added at rt under stirring in an argon atmosphere. An iodine solution (20 mM  $I_2$  in Py/ $H_2O$  = 9:1) was added to the reaction mixture until a slight yellow-brown color remained for 15 min (approx. 2.5 mL), followed by dd $H_2O$  (6.0 mL) and stirred for 1.5 h at rt. An aqueous solution of NaCl (3 M, 3.9 mL) was added, and the solution was transferred to a 50 mL centrifugation tube, vortexed vigorously and mixed with EtOH (absolute, 35 mL). Precipitation was achieved by cooling to  $-20^\circ C$  overnight, followed by centrifugation (10 min, 3200 x g). The supernatant was discarded and the resulting pellet was redissolved in buffer A (1.5 mL) and lyophilized. The residue was purified twice by RP-HPLC (0–20% B in 65 min). This afforded the benzaldehyde triphosphate **78** as a colorless tetrakis-triethylammonium salt (13.8 mg, 17.6  $\mu$ mol, 10%) after freeze-drying.

**$^1H$  NMR** (400 MHz,  $D_2O$ ):  $\delta$  (ppm) = 9.99 (s, 1H, 4''-H), 8.04 (d,  $^3J$  = 8.0 Hz, 2H, 2-H and 6-H), 7.77 (d,  $^3J$  = 8.7 Hz, 2H, 3-H and 5-H), 4.95 (d,  $^3J$  = 7.5 Hz, 1H, 1'-H), 4.42 (dd,  $J$  = 5.2, 3.2 Hz, 1H, 3'-H), 4.37–4.31 (m, 1H, 4'-H), 4.30–4.25 (m, 2H, 5'-H), 4.22 (dd,  $J$  = 7.5, 5.2 Hz, 1H, 2'-H).

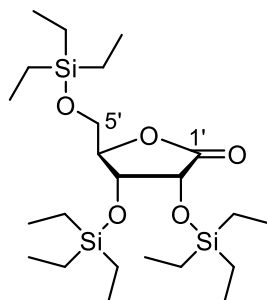
**$^{31}P$  NMR** (162 MHz,  $^1H$ -decoupled,  $D_2O$ ):  $\delta$  (ppm) =  $-10.74$  –  $-11.10$  (m, 1P,  $\gamma$ -P),  $-11.30$  (d,  $^2J_{PP}$  = 20.1 Hz, 1P,  $\alpha$ -P),  $-22.74$  –  $-23.50$  (m, 1P, P- $\beta$ ).

**HRMS (ESI):** calc. for  $C_{12}H_{15}Na_3O_{14}P_3^+ [M-2H+3Na]^+$ : 544.9362; found: 544.9362.

**Extinction coefficient** ( $H_2O$ ):  $\epsilon$  (260 nm) = 7200  $M^{-1}cm^{-1}$

### 5.2.4 Synthesis of a Salicylaldehyde Ribophosphoramidite

(3*R*,4*R*,5*R*)-3,4-bis(triethylsilyloxy)-5-((triethylsilyloxy)methyl)dihydrofuran-2(3*H*)-one (**83**)



$C_{23}H_{50}O_5Si_3$   
MW: 490.90 g/mol

$\beta$ -D-Ribose (3.0 g, 20 mmol, 1.0 eq) and  $NaHCO_3$  (3.4 g, 40 mmol, 2.0 eq) were dissolved in  $dH_2O$  (30 mL) and cooled to 0 °C. Bromine (1.1 mL, 22 mmol, 1.1 eq) was added dropwise to the stirring solution at 0 °C and the mixture was allowed to warm to rt slowly and stirred for 2 h. Excess bromine was quenched by titration with a solution of  $Na_2S_2O_3$  until complete color discharge and the solution was concentrated *in vacuo* (at 60 °C). The residue was excessively dried for at least 20 h under high vacuum. The resulting off-white solid containing ribolactone **85** was dissolved together with imidazol (6.8 g, 100 mmol, 5.0 eq) in dry DMF (50 mL) and cooled to 0 °C. Chlorotriethylsilane (13.5 mL, 80 mmol, 4.0 eq) was added dropwise under stirring and the solution was warmed to rt. After 3 h, the mixture was poured into  $dH_2O$  (20 mL) and was extracted with diethyl ether (3 x 50 mL). The combined organic layers were washed with saturated aqueous  $NaHCO_3$ ,  $NaCl$  and  $dH_2O$  (30 mL each), dried over  $Na_2SO_4$ , filtered and concentrated *in vacuo*. Flash column chromatography (silica, 20 x 10 cm, wet load, gradient; *i*Hex/EtOAc = 1:0  $\rightarrow$  100:1  $\rightarrow$  66:1  $\rightarrow$  49:1  $\rightarrow$  20:1) yielded the TES-protected ribolactone **83** (7.1 g, 14.4 mmol, 72%) as a colorless oil.

$R_f$  (*i*Hex/EtOAc, 9:1) = 0.6.

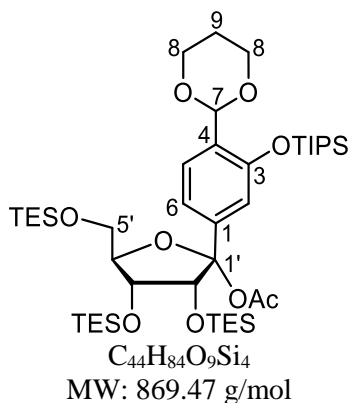
$^1H$  NMR (599 MHz,  $CDCl_3$ ):  $\delta$  (ppm) = 4.60 (d,  $J$  = 5.1 Hz, 1H, 2'-H), 4.30 (dd,  $J$  = 5.1, 0.8 Hz, 1H, 3'-H), 4.26 (t,  $J$  = 2.8 Hz, 1H, 4'-H), 3.83 (dd,  $J$  = 11.7, 3.2 Hz, 1H, 5'-H), 3.78 (dd,  $J$  = 11.6, 2.3, 1H, 5'-H), 1.02-0.89 (m, 27 H, 3 x  $Si(CH_2CH_3)_3$ ), 0.73-0.53 (m, 18 H, 3 x  $Si(CH_2CH_3)_3$ ).

$^{13}C$  NMR (151 MHz,  $^1H$ -decoupled,  $CDCl_3$ ):  $\delta$  (ppm) = 175.3 (1'-C), 85.8 (4'-H), 72.0 (3'-H), 70.4 (2'-H), 62.1 (5'-C), 6.7 (2 x  $Si(CH_2CH_3)_3$ ), 6.6 ( $Si(CH_2CH_3)_3$ ), 4.8 (2 x  $Si(CH_2CH_3)_3$ ), 4.1 ( $Si(CH_2CH_3)_3$ ).

IR (ATR):  $\tilde{\nu}$  ( $cm^{-1}$ ) = 2954, 2912, 2877, 1795, 1458, 1414, 1379, 1341, 1239, 1171, 1110, 1086, 1004, 984, 951, 899, 860, 801, 785, 721, 673.

MS (EI): calc. for  $C_{23}H_{50}O_5Si_3^+$   $[M]^+$ : 490.2960, found: 490.18.

(2*R*,3*R*,4*R*,5*R*)-2-(4-(1,3-Dioxan-2-yl)-3-(triisopropylsilyloxy)phenyl)-3,4-bis(triethylsilyloxy)-5-((triethylsilyloxy)methyl)tetrahydrofuran-2-yl acetate (**86**)



The protected salicyl base building block **2** (2.88 g, 6.93 mmol, 1.7 eq) was dissolved in dry diethyl ether (20 mL) under an argon atmosphere, degassed and cooled to  $-78^{\circ}\text{C}$ . Then, *t*BuLi (1.76 M in pentane, 7.88 mL, 13.86 mmol, 3.4 eq.) was added dropwise within 30 min and the solution was stirred for 2 h at  $-78^{\circ}\text{C}$ . TES-protected ribolactone **83** (2.00 g, 4.08 mmol, 1.0 eq) dissolved in dry THF (16 mL) was added dropwise. After 30 min, acetic anhydride (2.00 mL, 21.2 mmol, 5.2 eq) was added, and the reaction mixture was stirred for 15 min at  $-78^{\circ}\text{C}$  before it was allowed to warm to rt. The mixture was poured into ice water (120 mL) and was extracted with diethyl ether (3 x 100 mL). The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo* ( $30^{\circ}\text{C}$ ). Flash column chromatography (silica, 16 x 5.5 cm, wet load, gradient; *i*Hex/EtOAc = 100:1  $\rightarrow$  50:1  $\rightarrow$  33:1  $\rightarrow$  25:1  $\rightarrow$  20:1) afforded the acetylated hemiketal **86** (2.56 g, 2.94 mmol, 72%) as a colorless oil.

**R<sub>f</sub>** (*i*Hex/EtOAc, 9:1) = 0.4.

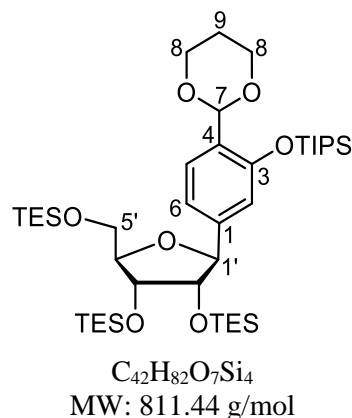
**<sup>1</sup>H NMR** (300 MHz, CDCl<sub>3</sub>):  $\delta$  (ppm) = 7.50 (d,  $^3J = 8.1$  Hz, 1H, 5-H), 7.16 (dd,  $^3J = 8.1$  Hz,  $^4J = 1.7$  Hz, 1H, 6-H), 6.82 (d,  $^3J = 1.7$  Hz, 1H, 2-H), 5.83 (s, 1H, 7-H), 4.29-4.14 (m, 4H, 4'-H, 3'-H and 8-H), 4.01-3.87 (m, 3H, 2'-H and 8-H), 3.82 (dd,  $J = 10.8, 3.4$  Hz, 1H, 5'-H), 3.75 (dd,  $J = 10.8, 4.9$  Hz, 1H, 5'-H), 2.34-2.12 (m, 1H, 9-H), 1.99 (s, 3H, OCOCH<sub>3</sub>), 1.36-1.21 (m, 4H, Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub> and 9-H), 1.12 (dd,  $J = 7.3, 3.8$  Hz, 18H, Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>), 1.02-0.92 (m, 18H, 2 x Si(CH<sub>2</sub>CH<sub>3</sub>)<sub>3</sub>), 0.92-0.82 (m, 9H, Si(CH<sub>2</sub>CH<sub>3</sub>)<sub>3</sub>), 0.68-0.56 (m, 12H, 2 x Si(CH<sub>2</sub>CH<sub>3</sub>)<sub>3</sub>), 0.53-0.43 (m, 6H, Si(CH<sub>2</sub>CH<sub>3</sub>)<sub>3</sub>).

**<sup>13</sup>C NMR** (75 MHz, <sup>1</sup>H-decoupled, CDCl<sub>3</sub>):  $\delta$  (ppm) = 168.1 (OCOCH<sub>3</sub>), 152.6 (1/3/4-C)\*, 142.0 (1/3/4-C)\*, 128.5 (5-C), 126.9 (1/3/4-C)\*, 118.4 (6-C), 115.3 (2-C), 106.3 (1'-C), 97.5 (7-C), 85.8 (4'-C), 80.2 (2'-C), 71.9 (3'-C), 67.5 (8-C), 67.4 (8-C), 62.4 (5'-C), 25.8 (9-C), 21.9 (OCOCH<sub>3</sub>), 18.0 (Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>), 13.0 (Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>), 6.8 (Si(CH<sub>2</sub>CH<sub>3</sub>)<sub>3</sub>), 6.70 (Si(CH<sub>2</sub>CH<sub>3</sub>)<sub>3</sub>), 6.67 (Si(CH<sub>2</sub>CH<sub>3</sub>)<sub>3</sub>), 5.0 (Si(CH<sub>2</sub>CH<sub>3</sub>)<sub>3</sub>), 4.8 (Si(CH<sub>2</sub>CH<sub>3</sub>)<sub>3</sub>), 4.2 (Si(CH<sub>2</sub>CH<sub>3</sub>)<sub>3</sub>).

**IR** (ATR):  $\tilde{\nu}$  (cm<sup>-1</sup>) = 2953, 2874, 2360, 2341, 1758, 1616, 1579, 1506, 1460, 1415, 1392, 1284, 1236, 1185, 1150, 1100, 1000, 948, 882, 859, 843, 826, 800, 778, 726, 682.

**HRMS (ESI)**: calc. for C<sub>42</sub>H<sub>81</sub>O<sub>7</sub>Si<sub>4</sub><sup>+</sup> [oxocarbenium from M]<sup>+</sup>: 809.5054; found: 809.5058.

((2*S*,3*S*,4*R*,5*R*)-2-(4-(1,3-Dioxan-2-yl)-3-(triisopropylsilyloxy)phenyl)-5-((triethylsilyloxy)methyl)tetrahydrofuran-3,4-diyl)bis(oxy)bis(triethylsilane) (**82**)



To a stirring solution of the acetylated hemiketal **86** (960 mg, 1.10 mmol, 1.0 eq) in dry toluene (8.5 mL) was added triethylsilane (0.53 mL, 3.31 mmol, 3.0 eq) at  $-30^{\circ}\text{C}$ . After 5 min stirring the solution was warmed to  $-20^{\circ}\text{C}$ ,  $\text{BF}_3\cdot\text{OEt}_2$  (0.15 mL, 1.21 mmol, 1.1 eq) was added dropwise and stirring was continued for 5 min. The deep dark green solution was quenched with an aqueous saturated solution of  $\text{NaHCO}_3$  (12.5 mL), diluted with ddH<sub>2</sub>O (12.5 mL) and warmed to rt. The mixture was immediately extracted with diethyl ether (3 x 50 mL) and the combined organic layers were dried over  $\text{Na}_2\text{SO}_4$ , filtered and dried *in vacuo*. Flash column chromatography (silica, 18 x 5.5 cm, wet load, gradient; *i*Hex/EtOAc = 1:0  $\rightarrow$  100:1  $\rightarrow$  50:1  $\rightarrow$  25:1) afforded the fully protected salicyl nucleoside **82** (322 mg, 0.40 mmol, 36%) as a colorless oil.

$R_f$  (*i*Hex/EtOAc, 9:1) = 0.5.

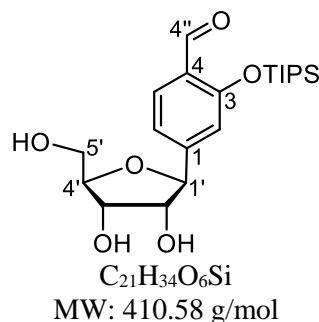
**<sup>1</sup>H NMR** (300 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 7.51 (d,  $J$  = 7.9 Hz, 1H, 5-H), 7.07 (dd,  $J$  = 8.1 Hz,  $J$  = 1.3 Hz, 1H, 6-H), 6.75 (d,  $J$  = 1.3 Hz, 1H, 2-H), 5.85 (s, 1H, 7-H), 4.73 (d,  $J$  = 5.7 Hz, 1H, 1'-H), 4.23 (dd,  $J$  = 11.6, 4.1 Hz, 2H, 8-H), 4.10 (t,  $J$  = 4.1 Hz, 1H, 3'-H), 4.05-3.99 (m, 1H, 4'-H), 3.99-3.89 (m, 2H, 8-H), 3.84 (dd,  $J$  = 5.6, 4.3 Hz, 1H, 2'-H), 3.80-3.67 (m, 2H, 5'-H), 2.36-2.12 (m, 1H, 9-H), 1.45-1.36 (m, 1H, 9-H), 1.37-1.24 (m, 3H,  $\text{Si}(\text{CH}(\text{CH}_3)_2)_3$ ), 1.13 (d,  $J$  = 7.0 Hz, 18H,  $\text{Si}(\text{CH}(\text{CH}_3)_2)_3$ ), 1.07-0.91 (m, 18H, 2 x  $\text{Si}(\text{CH}_2\text{CH}_3)_3$ ), 0.86 (t,  $J$  = 7.9 Hz, 9H,  $\text{Si}(\text{CH}_2\text{CH}_3)_3$ ), 0.71-0.57 (m, 12H, 2 x  $\text{Si}(\text{CH}_2\text{CH}_3)_3$ ), 0.53-0.39 (m, 6H,  $\text{Si}(\text{CH}_2\text{CH}_3)_3$ ).

**<sup>13</sup>C NMR** (75 MHz,  $^1\text{H}$ -decoupled,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 152.9 (1/3/4-C)\*, 142.7 (1/3/4-C)\*, 128.2 (1/3/4-C)\*, 127.0 (5-C), 119.4 (6-C), 116.0 (2-C), 97.6 (7-C), 84.1 (4'-C), 83.9 (1'-C), 79.4 (2'-C), 73.2 (3'-C), 67.6 (8-C), 63.3 (5'-C), 25.9 (9-C), 18.1 ( $\text{Si}(\text{CH}(\text{CH}_3)_2)_3$ ), 18.0 ( $\text{Si}(\text{CH}(\text{CH}_3)_2)_3$ ), 13.1 ( $\text{Si}(\text{CH}(\text{CH}_3)_2)_3$ ), 6.9 ( $\text{Si}(\text{CH}_2\text{CH}_3)_3$ ), 6.8 (2 x  $\text{Si}(\text{CH}_2\text{CH}_3)_3$ ), 5.0 ( $\text{Si}(\text{CH}_2\text{CH}_3)_3$ ), 4.8 ( $\text{Si}(\text{CH}_2\text{CH}_3)_3$ ), 4.3 ( $\text{Si}(\text{CH}_2\text{CH}_3)_3$ ).

**HRMS (ESI)**: calc. for  $\text{C}_{42}\text{H}_{83}\text{O}_7\text{Si}_4^+$  [ $\text{M}+\text{H}$ ] $^+$ : 811.5210; found: 811.5216.



4-((2*S*,3*R*,4*S*,5*R*)-3,4-Dihydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-2-(triisopropylsilyloxy)benzaldehyde (**87**)



**Route A:** To a stirring solution of the fully protected salicyl nucleoside **82** (249 mg, 0.31 mmol, 1.0 eq) in a mixture of THF/dH<sub>2</sub>O/AcOH (6:2:1, 4.5 mL) was added aqueous HCl (2 M, 0.3 mL). After stirring for 15 min at rt the mixture was poured into dH<sub>2</sub>O (10 mL) and extracted with diethyl ether (2 x 15 mL). The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo*. Flash column chromatography (silica, 15 x 2 cm, wet load, gradient; *i*Hex/EtOAc = 3:1 → 1:1 → 0:1) afforded the TIPS-protected salicylaldehyde nucleoside **87** (116 mg, 0.28 mmol, 91%) as a colorless oil.

**Route B:** To a stirring solution of the acetylated hemiketal **86** (2.27 g, 2.61 mmol, 1.0 eq) in dry toluene (17.1 mL) was added triethylsilane (1.25 mL, 7.83 mmol, 3.0 eq) at −30 °C. After 5 min stirring the solution was warmed to −20 °C, BF<sub>3</sub>·OEt<sub>2</sub> (0.17 mL, 1.31 mmol, 0.5 eq) was added dropwise and stirring was continued for 5 min. The deep dark green solution was quenched with an aqueous saturated solution of NaHCO<sub>3</sub> (13.0 mL), diluted with dH<sub>2</sub>O (13.0 mL) and warmed to rt. The mixture was immediately extracted with diethyl ether (3 x 60 mL) and the combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo*. The crude product mixture, which contained the fully protected nucleoside **82** (about 70% determined by <sup>1</sup>H NMR) was dissolved in a mixture of THF/dH<sub>2</sub>O/AcOH (6:2:1, 13.5 mL) and aqueous HCl (2 M, 1.2 mL). After stirring 1 h at rt, dH<sub>2</sub>O (50 mL) was added and the mixture was extracted with diethyl ether (3 x 50 mL). The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo*. Flash column chromatography (silica, 17 x 4 cm, wet load, gradient; *i*Hex/EtOAc = 3:1 → 1:1 → 0:1) afforded the TIPS-protected salicylaldehyde nucleoside **87** (649 mg, 1.58 mmol, 60% from acetylated hemiketal **86**) as a pale yellow oil.

*R<sub>f</sub>* (DCM) = 0.5.

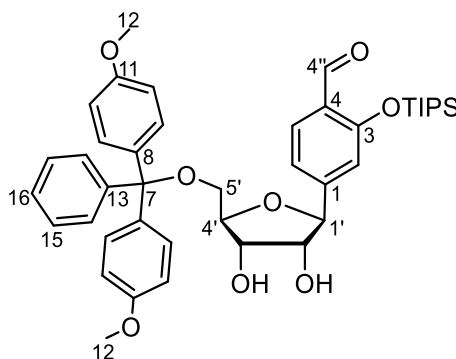
**<sup>1</sup>H NMR** (599 MHz, CDCl<sub>3</sub>): δ (ppm) = 10.48 (d, *J* = 0.8 Hz, 1H, 4''-H), 7.76 (d, *J* = 8.0 Hz, 1H, 5-H), 7.04 (d, *J* = 8.1 Hz, 1H, 6-H), 7.00-6.97 (m, 1H, 2-H), 4.72 (d, *J* = 6.6 Hz, 1H, 1'-H), 4.17-4.12 (m, 1H, 3'-H), 4.06 (q, *J* = 4.2, 1H, 4'-H), 3.93-3.86 (m, 2H, 1 x 5'-H and 2'-H), 3.84-3.76 (m, 1H, 5'-H), 1.34 (hept, *J* = 7.5 Hz, 3H, (Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>), 1.12 (d, *J* = 7.5 Hz, 18H, (Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>)).

**<sup>13</sup>C NMR** (151 MHz, <sup>1</sup>H-decoupled, CDCl<sub>3</sub>): δ (ppm) = 190.1 (4''-C), 159.6 (ipso-C), 148.5 (ipso-C), 128.5 (5-C), 126.1 (ipso-C), 118.2 (6-C), 116.8 (2-C), 84.5 (4'-C), 83.6 (1'-C), 77.4 (2'-C), 71.8 (3'-C), 63.0 (5'-C), 17.9 (Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>), 12.9 (Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>).

**IR** (ATR):  $\tilde{\nu}$  (cm<sup>-1</sup>) = 2952, 2912, 2875, 2360, 1616, 1581, 1508, 1460, 1426, 1393, 1273, 1237, 1214, 1150, 1100, 1044, 1001, 930, 882, 844, 819, 725, 681.

**HRMS (ESI)**: calc. for C<sub>21</sub>H<sub>35</sub>O<sub>6</sub>Si<sup>+</sup> [M+H]<sup>+</sup>: 411.2197; found: 411.2199.

4-((2*S*,3*R*,4*S*,5*R*)-5-((Bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-3,4-dihydroxytetrahydrofuran-2-yl)-2-(triisopropylsilyloxy)benzaldehyde (**89**)



C<sub>42</sub>H<sub>52</sub>O<sub>8</sub>Si  
MW: 712.94 g/mol

TIPS-protected salicylaldehyde nucleoside **87** (800 mg, 1.95 mmol, 1.0 eq) was coevaporated in dry pyridine (2 x 30 mL), dissolved in dry pyridine (30 mL) and stirred over molecular sieve (3 Å) for 14 h. 4,4'-Dimethoxytrityl chloride (706 mg, 2.08 mmol, 1.1 eq) was added to the nucleoside solution and stirred for 3 h at rt. Methanol was added (7.5 mL) and stirring was continued for 1 h before the solution was filtered and concentrated *in vacuo*. Flash column chromatography (deactivated silica, 20 x 5 cm, wet load, gradient; 1% pyridine in *i*Hex/EtOAc = 9:1 → 5:1 → 3:1) afforded the DMTr-TIPS-protected salicylaldehyde nucleoside **89** (947 mg, 1.33 mmol, 68%) as a colorless foam.

**R<sub>f</sub>** (*i*Hex/EtOAc, 1:1) = 0.7.

**mp.**: 46-50 °C.

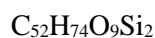
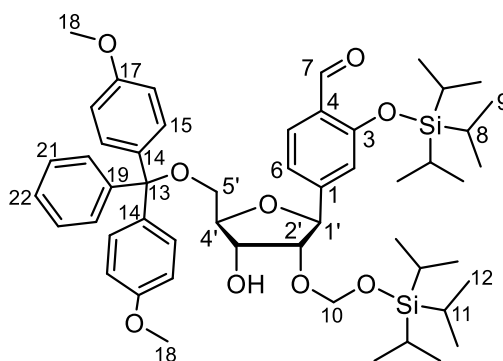
**<sup>1</sup>H NMR** (599 MHz, CDCl<sub>3</sub>):  $\delta$  = 10.51 (d, *J* = 0.7 Hz, 1H, 4''-H), 7.78 (d, *J* = 8.0 Hz, 1H, 5-H), 7.47-7.40 (m, 2H, 14-H), 7.35-7.31 (m, 4H, 9-H), 7.30-7.27 (m, 3H, 15-H, 16-H), 7.18-7.14 (m, 1H, 6-H), 6.96-6.93 (m, 1H, 2-H), 6.84-6.78 (m, 4H, 10-H), 4.73 (d, *J* = 6.6 Hz, 1H, 1'-H), 4.19-4.10 (m, 2H, 3'-H and 4'-H), 3.97 (dd, *J* = 6.6, 5.6 Hz, 1H, 2'-H), 3.79 (s, 6H, 12-H), 3.50-3.29 (m, 2H, 5'-H), 1.39-1.24 (m, 3H, Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>), 1.09 (dd, *J* = 7.5, 3.5 Hz, 18H, Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>).

**<sup>13</sup>C NMR** (151 MHz, <sup>1</sup>H-decoupled, CDCl<sub>3</sub>):  $\delta$  (ppm) = 189.9 (4''-C), 159.4 (3-C), 158.5 (2 x 11-C), 148.7 (1-C), 144.6 (13-C), 135.8 (2 x 8-C), 130.0 (4 x 9-C), 128.4 (5-C), 128.1 (2 x 14-C), 127.8 (2 x 15-C), 126.9 (6-C), 126.0 (4-C), 123.8 (16-C), 118.2 (6-C), 117.0 (2-C), 113.1 (4 x 10-C), 86.4 (7-C), 83.4 (1'-C, and 4'-C), 77.4 (2'-C), 72.9 (3'-C), 64.0 (5'-C), 55.2 (12-C), 17.9 (Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>), 12.9 (Si(CH(CH<sub>3</sub>)<sub>2</sub>)<sub>3</sub>).

**IR** (ATR):  $\tilde{\nu}$  (cm<sup>-1</sup>) = 3413, 3062, 2945, 2867, 2361, 2341, 2043, 1684, 1607, 1568, 1508, 1490, 1462, 1425, 1391, 1301, 1247, 1174, 1074, 1033, 996, 977, 915, 881, 855, 825, 790, 751, 726, 686, 661.

**HRMS (ESI)**: calc. for C<sub>42</sub>H<sub>51</sub>O<sub>8</sub>Si<sup>-</sup> [M-H]<sup>-</sup>: 711.3359; found: 711.3376.

4-((2*S*,3*R*,4*R*,5*R*)-5-((Bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-4-hydroxy-3-((triisopropylsilyloxy)methoxy)tetrahydrofuran-2-yl)-2-(triisopropylsilyloxy)benzaldehyde (**90a**)



MW: 899.31 g/mol

To a stirring solution of DMTr-TIPS salicylaldehyde nucleoside **89** (706 mg, 0.99 mmol, 1.0 eq) in dry 1,2-dichloroethane (4.75 mL) was added DIPEA (674  $\mu\text{L}$ , 3.96 mmol, 4.0 eq), di-*tert*-butyltin dichloride (365 mg, 1.20 mmol, 1.2 eq), and the reaction mixture was stirred at 70 °C for 15 min. After cooling to rt, (triisopropylsilyloxy)methyl chloride (251  $\mu\text{L}$ , 1.08 mmol, 1.1 eq) was slowly added dropwise and the mixture was stirred for 2 h. Methanol (940  $\mu\text{L}$ ) was added, the solution was diluted with DCM (20 mL) and washed with half-saturated aqueous  $\text{NaHCO}_3$  and  $\text{dH}_2\text{O}$  (10 mL each). The organic layer was dried over  $\text{Na}_2\text{SO}_4$ , filtered and concentrated *in vacuo*. Flash column chromatography (deactivated silica, 17 x 4 cm, wet load, gradient; 1% pyridine in *i*Hex/EtOAc = 19:1  $\rightarrow$  12:1  $\rightarrow$  10:1  $\rightarrow$  9:1) afforded a 1:1 regioisomer mixture (determined by  $^1\text{H}$  NMR) of 2'- and 3'-TOM-protected DMTr-TIPS salicylaldehyde nucleoside **90** (614 mg, 0.68 mmol, 68%). By painstaking repetitive flash column purification this mixture could be separated whereby the 3'-protected isomer **90b** (273 mg, 0.30 mmol, 30%) eluted first. This finally afforded the desired 2'-TOM DMTr TIPS salicylaldehyde nucleoside **90a** (278 mg, 0.31 mmol, 31%) as a colorless foam.

Data for the 2'-regioisomer **90a**:

$R_f$  (*i*Hex/EtOAc, 3:1) = 0.7.

mp.: 40-44 °C.

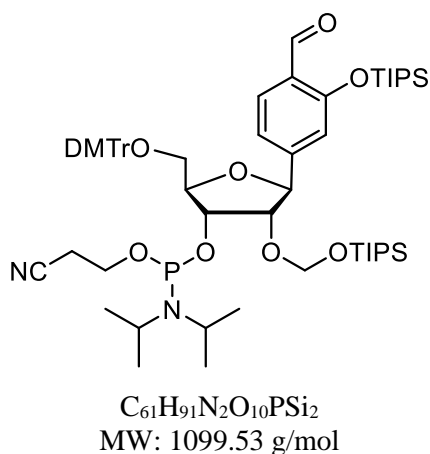
$^1\text{H}$  NMR (599 MHz,  $\text{CDCl}_3$ ):  $\delta$ (ppm) = 10.51 (d,  $J$  = 0.8 Hz, 1H, 7-H), 7.77 (d,  $J$  = 8.1 Hz, 1H, 5-H), 7.48-7.41 (m, 2H, 20-H), 7.37-7.30 (m, 4H, 15-H), 7.32-7.13 (m, 3H, 21+22-H), 7.23-7.18 (m, 1H, 6-H), 6.93-6.90 (m, 1H, 2-H), 6.83-6.79 (m, 4H, 16-H), 5.14 (d,  $J$  = 4.6 Hz, 1H, 10-H), 4.86 (d,  $^3J$  = 7.8 Hz, 1H, 1'-H), 4.78 (d,  $J$  = 4.7 Hz, 1H, 10-H), 4.30-4.26 (m, 1H, 3'-H), 4.26-4.21 (m, 1H, 4'-H), 3.80-3.79 (m, 1H, 2'-H), 3.79 (s, 6H, 18-H), 3.37 (dd,  $J$  = 10.0, 4.0 Hz, 1H, 5'-H), 3.33 (dd,  $J$  = 10.0, 4.4 Hz, 1H, 5'-H), 3.06 (d,  $^3J$  = 2.0 Hz, 1H, 3'-OH), 1.39-1.21 (m, 6H, 11+8-H), 1.10 (d,  $^3J$  = 7.4 Hz, 18H, 9-H), 1.08-1.04 (m, 18H, 12-H).

**$^{13}\text{C}$  NMR** (151 MHz,  $^1\text{H}$ -decoupled,  $\text{CDCl}_3$ ):  $\delta$ (ppm) = 190.0 (7-C), 159.4 (3-C), 158.6 (2 x 17-C), 149.1 (1-C), 144.9 (19-C), 136.2 (14-C), 136.1 (14-C), 130.2 (4 x 15-C), 128.6 (5-C), 128.4 (2 x 20-C), 127.9 (2 x 21-C), 126.9 (22-C), 126.2 (4-C), 118.4 (6-C), 117.4 (2-C), 113.3 (4 x 16-C), 91.3 (10-C), 87.0 (2'-C), 86.4 (13-C), 84.4 (4'-C), 80.8 (1'-C), 72.6 (3'-C), 64.2 (5'-C), 55.4 (2 x 18-C), 18.1 (6 x 9-C), 18.0 (6 x 12-C), 13.1 (3 x 8-C), 12.0 (3 x 11-C).

**IR** (ATR):  $\tilde{\nu}$  ( $\text{cm}^{-1}$ ) = 3526, 2944, 2892, 2866, 2360, 2341, 1686, 1608, 1569, 1508, 1490, 1462, 1424, 1390, 1368, 1300, 1248, 1174, 1075, 1034, 1012, 993, 978, 917, 881, 825, 790, 772, 752, 725, 683, 661.

**HRMS (ESI)**: calc. for  $\text{C}_{52}\text{H}_{75}\text{O}_9\text{Si}_2^+$   $[\text{M}+\text{H}]^+$ : 899.4944; found: 899.4955.

(2*R*,3*R*,4*S*,5*S*)-2-((Bis(4-methoxyphenyl)(phenyl)methoxy)methyl)-5-(4-formyl-3-(triisopropylsilyloxy)phenyl)-4-((triisopropylsilyloxy)methoxy)tetrahydrofuran-3-yl 2-cyanoethyl diisopropylphosphoramidite (**88**)

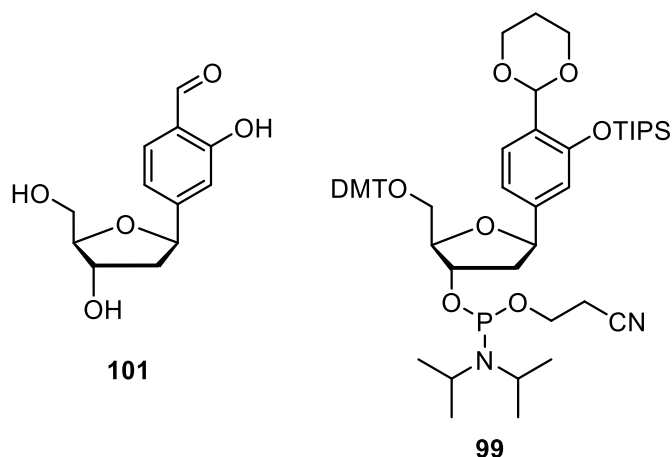


A solution of 2'-TOM-DMTr-TIPS salicylaldehyde nucleoside **90a** (100 mg, 111  $\mu\text{mol}$ , 1.0 eq) and DIPEA (59  $\mu\text{L}$ , 45 mg, 345  $\mu\text{mol}$ , 3.1 eq) in dry DCM (0.52 mL) was degassed (3x) and kept under an atmosphere of argon. 2-Cyanoethyl *N,N*-diisopropylchlorophosphoramidite (38  $\mu\text{L}$ , 40 mg, 167  $\mu\text{mol}$ , 1.5 eq) was added under stirring at rt. After 2 h, the turbid mixture was directly applied to flash column chromatography (deactivated silica, argon, 10 x 2 cm, 1% pyridine in *i*Hex/EtOAc = 6:1) and quickly purified to afford the salicylaldehyde ribophosphoramidite **88** (109 mg, 99  $\mu\text{mol}$ , 89%) as a colorless foam.

**$^{31}\text{P}$  NMR** (81 MHz,  $^1\text{H}$ -decoupled,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 151.3, 151.0.

**HRMS (ESI)**: calc. for  $\text{C}_{61}\text{H}_{92}\text{N}_2\text{O}_{10}\text{PSi}_2^+$   $[\text{M}+\text{H}]^+$ : 1099.6023; found: 1099.6023.

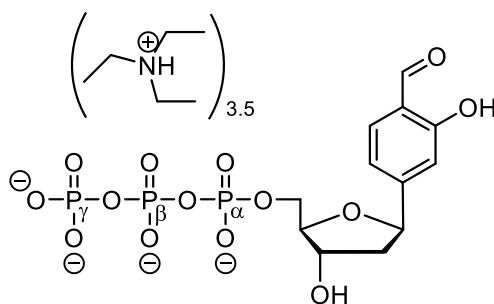
### 5.2.5 Synthesis of the Salicylaldehyde Nucleoside and Phosphoramidite



Salicylaldehyde nucleoside **101** and its phosphoramidite **99** were prepared according to the procedures reported in *G. H. Clever et al.* and *C. Kaul et al.* [12-13]

### 5.2.6 Synthesis of dSTP

[(2*R*,3*S*,5*R*)-5-(4-Formyl-3-hydroxyphenyl)-3-hydroxytetrahydrofuran-2-yl]methyl triphosphate (**102**)



C<sub>12</sub>H<sub>17</sub>O<sub>14</sub>P<sub>3</sub> (hypothetic free acid)  
MW: 478.18 g/mol

The synthesis of dSTP was performed according to the method described in *J. Caton-Williams et al.* from salicylaldehyde nucleoside **101** with only minor adjustments.<sup>[7]</sup> The analytical data are in agreement with *C. Kaul et al.* with only minor chemical shift changes due to the salt form and the different external references.<sup>[8]</sup>

In an argon atmosphere, triphosphate reagent solution (1.77 mL, 336 μmol, 2.0 eq, prepared according to general procedure **1**, p. 100) was added to the dry salicylaldehyde nucleoside **101** (40 mg, 168 μmol, 1.0 eq) at rt and stirred for 1 h. After complete conversion (monitored by TLC, DCM/MeOH = 5:1), an iodine solution (20 mM I<sub>2</sub> in Py/H<sub>2</sub>O = 9:1) was added until a permanent brown color was observed (approx. 1.6 mL) and stirring was continued for 15 min. ddH<sub>2</sub>O (4.1 mL) was added and after 1.5 h stirring, the resulting solution was transferred to a 50 mL centrifuge tube. Brine was added to a final concentration of 0.3 M, followed by addition of EtOH (absolute, 36 mL) and vortexing. After 12 h at -20 °C, the precipitate was collected by centrifugation (10 min, 3200 x g) and the resulting pellet was purified twice by

preparative RP-HPLC (0–20% B in 45 min). This afforded the dSTP **102** as a yellow tris- to tetrakis-triethylammonium salt (21 mg, 26  $\mu$ mol, 15%) after freeze-drying.

**$^1\text{H}$  NMR** (400 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = 9.94 (s, 1H,  $\text{H}_{\text{aldehyde}}$ ), 7.73 (d,  $J = 8.1$  Hz, 1H,  $\text{H}_{\text{ar}}$ ), 7.14 (dd,  $J = 8.1$  Hz,  $^4J = 1.5$  Hz, 1H,  $\text{H}_{\text{ar}}$ ), 7.10 (d,  $J = 1.5$  Hz, 1H,  $\text{H}_{\text{ar}}$ ), 5.21 (dd,  $J = 10.5$ ,  $J = 6.0$  Hz, 1H,  $1'\text{-H}$ ), 4.55 (d, 1H,  $J = 5.6$  Hz, 1H,  $3'\text{-H}$ ), 4.28–4.17 (m, 1H,  $4'\text{-H}$ ), 4.18–4.03 (m, 2H,  $5'\text{-H}$ ), 2.30 (ddd,  $J = 13.5$  Hz,  $J = 5.7$  Hz,  $J = 1.3$  Hz, 1H,  $2'\text{-H}$ ), 2.10 (ddd,  $J = 13.6$  Hz,  $J = 10.6$  Hz,  $J = 5.7$ , 1H,  $2'\text{-H}$ ).

**$^{31}\text{P}$  NMR** (162 MHz,  $^1\text{H}$ -decoupled,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = –11.02 (d,  $^2J_{\text{PP}} = 19.8$  Hz, 1P,  $\gamma\text{-P}$ ), –11.31 (d,  $^2J_{\text{PP}} = 20.2$  Hz, 1P,  $\alpha\text{-P}$ ), –23.47 (t,  $^2J_{\text{PP}} = 20.0$  Hz, 1P,  $\beta\text{-P}$ ).

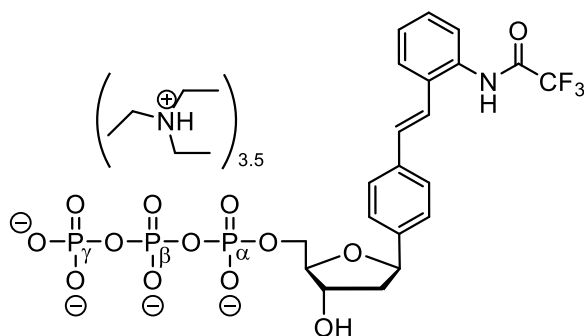
**HRMS (ESI)**: calc. for  $\text{C}_{12}\text{H}_{16}\text{O}_{14}\text{P}_3^- [\text{M-H}]^-$ : 476.9758; found: 476.9767.

**UV-Vis** ( $\text{H}_2\text{O}$ ):  $\lambda_{\text{Abs}}$  (nm) = 230, 290, 330.

**Extinction coefficient** ( $\text{H}_2\text{O}$ ):  $\epsilon$  (260 nm) = 10200  $\text{M}^{-1}\text{cm}^{-1}$ .

### 5.2.7 Synthesis of dT<sub>0</sub>TP

{(2*R*,3*S*,5*R*)-3-Hydroxy-5-{4-[2-(2,2,2-trifluoroacetamido)styryl]phenyl}tetrahydrofuran-2-yl}methyl triphosphate (**106**)



$\text{C}_{21}\text{H}_{23}\text{F}_3\text{NO}_{13}\text{P}_3$  (hypothetic free acid)  
MW: 647.33 g/mol

In an argon atmosphere, triphosphate reagent solution (2.77 mL, 526  $\mu$ mol, 2.0 eq, prepared according to general procedure **1**, p. 100) was added to the dry TFA-protected aniline nucleoside **104** (107 mg, 263  $\mu$ mol, 1.0 eq) at 0 °C and stirred for 3 h. An iodine solution (20 mM  $\text{I}_2$  in  $\text{Py}/\text{H}_2\text{O} = 9:1$ ) was added until a permanent brown color was observed (approx. 2.5 mL) and stirring was continued for 15 min.  $\text{ddH}_2\text{O}$  (6.4 mL) was added and after 1.5 h stirring, the resulting solution was partitioned to two 50 mL centrifuge tubes. Brine was added to a final concentration of 0.3 M followed by the addition of EtOH (absolute, 24 mL) each and vortexing. After 1 h at –80 °C, the precipitate was collected by centrifugation (10 min, 3200 x g) and the resulting pellet was purified twice by preparative RP-HPLC (0–50% B in 45 min). This afforded the TFA-protected dT<sub>0</sub>TP **106** as a yellow fluorescent tris- to tetrakis-triethylammonium salt (63 mg, 63  $\mu$ mol, 24%) after freeze-drying.

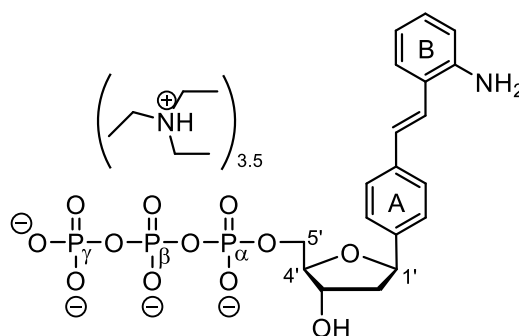
**$^1\text{H}$  NMR** (400 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = 7.71 (dd,  $J = 7.5$  Hz,  $J = 1.9$  Hz, 1H,  $\text{H}_{\text{ar}}$ ), 7.46 (d,  $J = 8.3$  Hz, 2H,  $\text{H}_{\text{ar}}$ ), 7.40–7.31 (m, 4H,  $\text{H}_{\text{ar}}$ ), 7.27 (dd,  $J = 7.5$  Hz,  $J = 1.8$  Hz, 1H,  $\text{H}_{\text{ar}}$ ), 7.11 (d,  $J = 16.3$  Hz, 1H,  $\text{CH}=\text{CH}$ ),

7.03 (d,  $J = 16.4$  Hz, 1H,  $CH=CH$ ), 5.09 (dd,  $J = 10.5$  Hz,  $J = 5.5$  Hz, 1H,  $1'-H$ ), 4.60–4.44 (m, 1H,  $3'-H$ ), 4.18–4.11 (m, 1H,  $4'-H$ ), 4.09–4.00 (m, 2H,  $5'-H$ ), 2.22–2.13 (m, 1H,  $2'-H$ ), 2.12–2.01 (m, 1H,  $2'-H$ ).

$^{31}\text{P}$  NMR (162 MHz,  $^1\text{H}$ -decoupled,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) =  $-11.02$  (d, 1P,  $^2J_{\text{PP}} = 19.8$  Hz,  $\gamma/\alpha\text{-P}$ ),  $-11.36$  (d, 1P,  $^2J_{\text{PP}} = 20.1$  Hz,  $\gamma/\alpha\text{-P}$ ),  $-23.53$  (t, 1P,  $^2J_{\text{PP}} = 18.8$  Hz,  $\beta\text{-P}$ ).

HRMS (ESI): calc. for  $\text{C}_{21}\text{H}_{22}\text{O}_{13}\text{NF}_3\text{P}_3^-$  [ $\text{M}-\text{H}$ ] $^-$ : 646.0262; found: 646.0253.

{(2*R*,3*S*,5*R*)-5-[4-(2-Aminostyryl)phenyl]-3-hydroxytetrahydrofuran-2-yl}methyl triphosphate (**105**)



$\text{C}_{19}\text{H}_{24}\text{NO}_{12}\text{P}_3$  (hypothetic free acid)  
MW: 551.32 g/mol

**Route A:** To a freshly prepared aqueous solution of  $\text{NH}_4\text{OH}$  (28%) and  $\text{NH}_2\text{Me}$  (40%) (50:50, v/v, 3 mL) was added the triethylammonium salt of TFA-protected dT<sub>o</sub>TP **106** (63 mg, 63  $\mu\text{mol}$ ) and heated to 65 °C. For reaction control, aliquots were taken for analytical RP-HPLC (0–70% B in 45 min) in 30 min intervals. After 2 h almost complete conversion of the starting material was observed. After cooling to 0 °C, acetic acid (approx. 1.8 mL) was added to the yellow solution for neutralization and the filtered sample was purified twice by preparative RP-HPLC (0–40% B in 45 min). This afforded the dT<sub>o</sub>TP **105** as a yellow fluorescent tris- to tetrakis-triethylammonium salt (22.2 mg, 24.6  $\mu\text{mol}$ , 39%) after freeze-drying.

**Route B:** To the excessively dried aromatic amine nucleoside **103** (18.1 mg, 58.0  $\mu\text{mol}$ , 1.0 eq) was added the triphosphate reagent solution (0.61 mL, 116.0  $\mu\text{mol}$ , 2.0 eq, prepared according to general procedure **1**, p. 100) at 0 °C in an argon atmosphere. The reaction mixture was allowed to warm to rt under stirring. After 3.0 h, an iodine solution (20 mM  $\text{I}_2$  in  $\text{Py}/\text{H}_2\text{O} = 9:1$ ) was added until a slight brown color remained for 15 min (approx. 0.6 mL), followed by addition of ddH<sub>2</sub>O (1.5 mL) and stirring for 1.5 h. An aqueous solution of NaCl (3 M, 0.5 mL) was added, and the solution was transferred to a 50 mL centrifugation tube, vortexed vigorously and mixed with EtOH (absolute, 15 mL). Precipitation was achieved by cooling to –80 °C for 30 min, followed by centrifugation (5 min, 3200 x g). The supernatant was discarded and the pellet was redissolved in buffer A (1.0 mL) and lyophilized. The residue was purified twice by RP-HPLC purification (0–40% B in 45 min). This afforded the aromatic amine triphosphate **105** as a yellow fluorescent tetrakis-triethylammonium salt (4.7 mg, 4.9  $\mu\text{mol}$ , 8%) after freeze-drying.

$^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = 7.55 (d,  $J = 8.2$  Hz, 2H,  $\text{H}_{\text{ar-A}}$ ), 7.48 (d,  $J = 7.4$  Hz, 1H,  $\text{H}_{\text{ar-B}}$ ), 7.41 (d,  $J = 8.2$  Hz, 2H,  $\text{H}_{\text{ar-A}}$ ), 7.26 (d,  $J = 16.2$  Hz, 1H,  $CH=CH$ ), 7.14 (td,  $J = 7.8$  Hz, 1.2 Hz, 1H,  $\text{H}_{\text{ar-B}}$ ), 7.01

(d,  $J = 16.2$  Hz, 1H, CH=CH), 6.86 (t,  $J = 7.5$  Hz, 2H, H<sub>ar-B</sub>), 5.13 (dd,  $J = 10.5$  Hz,  $^3J = 5.6$  Hz, 1H, 1'-H), 4.59–4.42 (m, 1H, 3'-H), 4.18–4.13 (m, 1H, 4'-H), 4.12–4.05 (m, 2H, 5'-H), 2.24–2.17 (m, 1H, 2'-H), 2.16–2.06 (m, 1H, 2'-H).

**$^{13}\text{C}$  NMR** (101 MHz,  $^1\text{H}$ -decoupled, D<sub>2</sub>O):  $\delta$  (ppm) = 143.1 (C<sub>ar</sub>), 139.9 (C<sub>ar</sub>), 137.2 (C<sub>ar</sub>), 129.2 (CH=CH), 129.0 (CH<sub>ar-B</sub>), 127.1 (2 x CH<sub>ar-A</sub>), 126.6 (2 x CH<sub>ar-A</sub>), 126.3 (CH<sub>ar-B</sub>), 124.3 (C<sub>ar</sub>), 123.4 (CH=CH), 120.3 (CH<sub>ar-B</sub>), 117.7 (CH<sub>ar-B</sub>), 85.6 (d,  $^3J_{\text{PC}} = 8.9$  Hz, 4'-C), 80.6 (1'-C), 73.0 (3'-C), 66.1 (d,  $^2J_{\text{PC}} = 5.9$  Hz, 5'-C), 41.8 (2'-C).

**$^{31}\text{P}$  NMR** (162 MHz,  $^1\text{H}$ -decoupled, D<sub>2</sub>O):  $\delta$  (ppm) = −10.80 (d, 1P,  $^2J_{\text{PP}} = 19.9$  Hz,  $\gamma$ -P), −11.32 (d, 1P,  $^2J_{\text{PP}} = 20.2$  Hz,  $\alpha$ -P), −23.41 (t, 1P,  $^2J_{\text{PP}} = 20.0$  Hz,  $\beta$ -P).

**HRMS (ESI)**: calc. for C<sub>19</sub>H<sub>23</sub>O<sub>12</sub>NP<sub>3</sub><sup>−</sup> [M−H]<sup>−</sup>: 550.0439; found: 550.0432.

**UV-Vis** (H<sub>2</sub>O):  $\lambda_{\text{Abs}}$  (nm) = 289 (base peak), 233, 330.

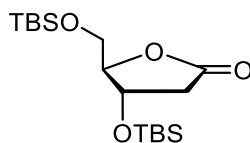
**Extinction coefficient** (H<sub>2</sub>O):  $\epsilon$  (289 nm) = 13500 M<sup>−1</sup>cm<sup>−1</sup>.

**Fluorescence** (H<sub>2</sub>O):  $\lambda_{\text{Abs}} = 289$  nm,  $\lambda_{\text{Em}} = 481$  nm.



### 5.2.8 Synthesis of a Naphthalene Amine Triphosphate

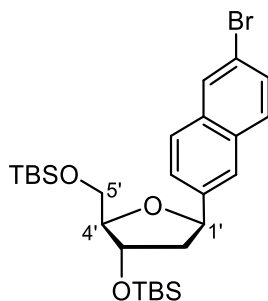
(4*S*,5*R*)-4-(*tert*-Butyldimethylsilyloxy)-5-((*tert*-butyldimethylsilyloxy)methyl)dihydrofuran-2(3*H*)-one (**91**)



$C_{17}H_{36}O_4Si_2$   
MW: 360.64 g/mol

The synthesis and the analytical data were in agreement with published data by Y. Cen *et al.*<sup>[224]</sup>

((2*R*,3*S*)-5-(6-Bromonaphthalen-2-yl)-2-((*tert*-butyldimethylsilyloxy)methyl)tetrahydrofuran-3-yloxy)(*tert*-butyl)dimethylsilane (**108**)



$C_{27}H_{43}BrO_3Si_2$   
MW: 551.70 g/mol

To a stirring solution of 2,6-dibromo-naphthalene (1.00 g, 3.50 mmol, 1.7 eq) in dry THF (13.5 mL) at  $-78\text{ }^{\circ}\text{C}$ , *n*BuLi (1.6 M in hexanes, 2.19 mL, 3.50 mmol, 1.7 eq) was added in one shot. After 20 min, a solution of the TBS-protected deoxyribolactone **91** (757 mg, 2.10 mmol, 1.0 eq) in dry THF (13.5 mL) was added dropwise to the yellow solution at  $-78\text{ }^{\circ}\text{C}$ , and stirring continued for 1 h. Then, the solution was poured into an ice-cold saturated aqueous  $NH_4Cl$  solution (60 mL) and was subsequently extracted with diethyl ether (3 x 70 mL). The combined organic layers were dried over  $Na_2SO_4$ , filtered and concentrated *in vacuo*. Flash column chromatography (silica, 20 x 4 cm, wet load, gradient; *i*Hex/EtOAc = 100:1  $\rightarrow$  50:1  $\rightarrow$  25:1  $\rightarrow$  10:1) yielded the hemiketal **109** as a colorless oil (271 mg, 0.48 mmol, 22%) that was directly used for the subsequent reduction.

$R_f$  (*i*Hex/EtOAc, 9:1) = 0.4.

To a stirring solution of the hemiketal **109** (270 mg, 0.48 mmol, 1.0 eq) in dry DCM (3 mL) at  $-78\text{ }^{\circ}\text{C}$ , triethylsilane (0.23 mL, 168 mg, 1.43 mmol, 3.0 eq) was added and after 5 min was followed by  $BF_3 \cdot OEt_2$  (73  $\mu$ L, 83 mg, 0.57 mmol, 1.2 eq). After 2 h, saturated aqueous  $NaHCO_3$  (3 mL),  $H_2O$  (10 mL) and DCM (7 mL) were added and the organic layer was separated after warming to rt. The aqueous layer was extracted with DCM (3 x 30 mL) and the combined organic layers were dried over  $Na_2SO_4$ , filtered and concentrated *in vacuo*. Flash column chromatography (silica, 18 x 2 cm, wet load, gradient; *i*Hex/EtOAc =

1:0 → 100:1 → 50:1 → 25:1) yielded the bromonaphthalene nucleoside **108** (134 mg, 0.24 mmol, 50% from **109**) as a colorless oil.

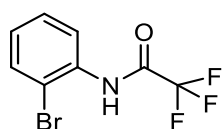
$R_f$  (*i*Hex/EtOAc, 9:1) = 0.6.

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>):  $\delta$  (ppm) = 7.97 (s, 1H, H<sub>ar</sub>), 7.80 (s, 1H, H<sub>ar</sub>), 7.71 (d,  $J$  = 8.7 Hz, 1H, H<sub>ar</sub>), 7.69 (d,  $J$  = 8.7 Hz, 1H, H<sub>ar</sub>), 7.53 (d,  $J$  = 8.6 Hz, 2H, H<sub>ar</sub>), 5.29 (dd,  $J$  = 10.5, 5.3 Hz, 1H, 1'-H), 4.46 (d,  $J$  = 4.9 Hz, 1H, 3'-H), 4.08–3.96 (s, 1H, 4'-H), 3.82 (dd,  $J$  = 10.7, 3.6 Hz, 1H, 5'-H), 3.68 (dd,  $J$  = 10.7, 5.6 Hz, 1H, 5'-H), 2.18 (dd,  $J$  = 12.5, 5.5 Hz, 1H, 2'-H), 1.96 (td,  $J$  = 12.2, 11.7, 5.3 Hz, 1H, 2'-H), 1.05–0.78 (m, 18H, 2 x SiC(CH<sub>3</sub>)<sub>3</sub>), 0.13–0.08 (m, 12H, 2 x Si(CH<sub>3</sub>)<sub>2</sub>).

**<sup>13</sup>C NMR** (101 MHz, <sup>1</sup>H-decoupled, CDCl<sub>3</sub>):  $\delta$  (ppm) = 140.4 (C<sub>ar</sub>), 134.0 (C<sub>ar</sub>), 131.7 (C<sub>ar</sub>), 129.7 (HC<sub>ar</sub>), 129.5 (HC<sub>ar</sub>), 129.3 (HC<sub>ar</sub>), 127.1 (HC<sub>ar</sub>), 125.4 (HC<sub>ar</sub>), 124.6 (HC<sub>ar</sub>), 119.5 (C<sub>ar</sub>), 88.2 (4'-C), 80.1 (1'-C), 74.5 (3'-C), 63.9 (5'-C), 44.4 (2'-C), 25.9 (SiC(CH<sub>3</sub>)<sub>3</sub>), 25.8 (SiC(CH<sub>3</sub>)<sub>3</sub>), 18.4 (SiC(CH<sub>3</sub>)<sub>3</sub>), 18.1 (SiC(CH<sub>3</sub>)<sub>3</sub>), -4.6 (SiCH<sub>3</sub>), -4.7 (SiCH<sub>3</sub>), -5.3 (SiCH<sub>3</sub>), -5.4 (SiCH<sub>3</sub>).

**HRMS (ESI)**: calc. for C<sub>27</sub>H<sub>47</sub>O<sub>3</sub>N<sup>81</sup>BrSi<sub>2</sub><sup>+</sup> [M+NH<sub>4</sub>]<sup>+</sup>: 570.22574; found: 570.22614.

*N*-(2-Bromophenyl)-2,2,2-trifluoroacetamide



C<sub>8</sub>H<sub>5</sub>BrF<sub>3</sub>NO  
MW: 268.03 g/mol

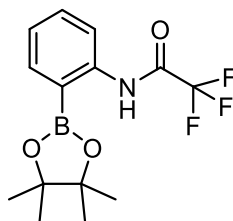
To a stirring solution of 2-bromoaniline (6.60 g, 38.4 mmol, 1.0 eq) in anhydrous DCM (75 mL) at 0 °C, triethylamine (16.0 mL, 11.65 g, 115.1 mmol, 3.0 eq) was added. After 5 min, trifluoroacetic anhydride (16.1 mL, 24.17 g, 115.1 mmol, 3.0 eq) was added dropwise and the reaction mixture was allowed to warm to rt. After 2 h, the mixture was poured into H<sub>2</sub>O (200 mL) and was extracted with DCM (3 x 150 mL). The combined organic layers were washed with H<sub>2</sub>O (1 x 200 mL), saturated aqueous NaHCO<sub>3</sub> (1 x 200 mL), dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo*. Flash column chromatography (silica, 17 x 6.0 cm, wet load, gradient; *i*Hex/EtOAc = 1:0 → 100:1 → 50:1 → 25:1) yielded the TFA-protected 2-bromoaniline (10.1 g, 37.7 mmol, 98% yield) as a colorless solid.

$R_f$  (*i*Hex/EtOAc, 9:1) = 0.4.

**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>):  $\delta$  (ppm) = 8.46 (br s, 1H, -NH), 8.32 (d,  $J$  = 8.2 Hz, 1H, H<sub>ar</sub>), 7.61 (d,  $J$  = 8.1 Hz, 1H, H<sub>ar</sub>), 7.40 (t,  $J$  = 7.8 Hz, 1H, H<sub>ar</sub>), 7.13 (t,  $J$  = 7.8 Hz, 1H, H<sub>ar</sub>).

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>):  $\delta$  (ppm) = 154.8 (q, <sup>2</sup> $J_{C-F}$  = 37 Hz, CO), 133.3 (C<sub>ar</sub>), 132.8 (HC<sub>ar</sub>), 128.9 (HC<sub>ar</sub>), 127.4 (HC<sub>ar</sub>), 122.2 (HC<sub>ar</sub>), 115.7 (q, <sup>1</sup> $J_{C-F}$  = 288 Hz, CF<sub>3</sub>), 114.2 (C<sub>ar</sub>).

2,2,2-Trifluoro-*N*-(2-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)phenyl)acetamide (**110**)



$\text{C}_{14}\text{H}_{17}\text{BF}_3\text{NO}_3$   
MW: 315.10 g/mol

To a stirring mixture of TFA-protected 2-bromoaniline (422 mg, 1.58 mmol, 1.0 eq) in dry dioxane (4.0 mL) in a 25 mL sealed tube bis(pinacolato)diboron (800 mg, 3.15 mmol, 2.0 eq) and potassium acetate (464 mg, 4.73 mmol, 3.0 eq) were added and the mixture was flushed with argon for 5 min. Then [1,1'-bis(diphenylphosphino)ferrocene]dichloropalladium(II) (57.6 mg, 78.8  $\mu\text{mol}$ , 5 mol%) was added and the mixture was heated to 95 °C for 2 h. After cooling to rt  $\text{H}_2\text{O}$  and DCM were added (90 mL each) and the organic layer was separated. The aqueous layer was extracted with DCM (3 x 90 mL) and the combined organic layers were washed with brine (90 mL), dried over  $\text{Na}_2\text{SO}_4$ , filtered and concentrated *in vacuo*. Flash column chromatography (silica, 20 x 2.5 cm, wet load, gradient; *i*Hex/EtOAc = 1:0  $\rightarrow$  100:1  $\rightarrow$  50:1) yielded the boronate amine **110** (422 mg, 1.34 mmol, 85% yield) as an off-white wax.

$R_f$  (*i*Hex/EtOAc, 9:1) = 0.6.

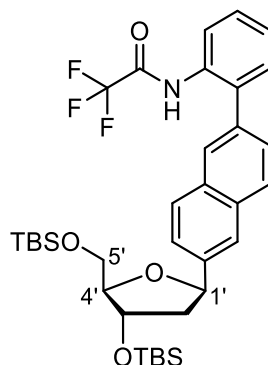
$^1\text{H NMR}$  (800 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 10.57 (s, 1H, -NH), 8.46 (d,  $J$  = 8.4 Hz, 1H,  $\text{H}_{\text{ar}}$ ), 7.81 (dd,  $J$  = 7.4, 1.6 Hz, 1H,  $\text{H}_{\text{ar}}$ ), 7.53 (td,  $J$  = 7.9, 1.6 Hz 1H,  $\text{H}_{\text{ar}}$ ), 7.20 (td,  $J$  = 7.4, 0.9 Hz, 1H,  $\text{H}_{\text{ar}}$ ), 1.38 (s, 12H, O-C-( $\text{CH}_3$ )<sub>2</sub>).

$^{13}\text{C NMR}$  (201 MHz,  $^1\text{H}$ -decoupled,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = 154.8 (q,  $^2J_{\text{C-F}}$  = 37 Hz, CO), 142.4 ( $\text{C}_{\text{ar}}$ ), 136.5 ( $\text{HC}_{\text{ar}}$ ), 133.3 ( $\text{HC}_{\text{ar}}$ ), 125.0 ( $\text{HC}_{\text{ar}}$ ), 119.7 ( $\text{HC}_{\text{ar}}$ ), 116.1 (q,  $^1J_{\text{C-F}}$  = 287 Hz,  $\text{CF}_3$ ), 84.9 (2 x OC( $\text{CH}_3$ )<sub>2</sub>), 24.8 (4 x C( $\text{CH}_3$ )<sub>2</sub>). The carbon directly adjacent to the boron residue was not detected.

$^{19}\text{F NMR}$  (376 MHz,  $\text{CDCl}_3$ ):  $\delta$  (ppm) = -76.3 (s, 3F).

**HRMS (ESI)**: calc. for  $\text{C}_{14}\text{H}_{16}\text{BF}_3\text{NO}_3^-$  [ $\text{M-H}$ ] $^-$ : 314.11808; found: 314.11825.

*N*-(2-(6-((4*S*,5*R*)-4-(*tert*-Butyldimethylsilyloxy)-5-((*tert*-butyldimethylsilyloxy)methyl)-tetrahydrofuran-2-yl)naphthalen-2-yl)phenyl)-2,2,2-trifluoroacetamide (**111**)



$C_{35}H_{48}F_3NO_4Si_2$   
MW: 659.93 g/mol

To a stirring solution of the bromonaphthalene nucleoside **108** (197 mg, 0.36 mmol, 1.0 eq) in DMF/H<sub>2</sub>O (9:1, 2 mL) the boronate amine **110** (170 mg, 0.54 mmol, 1.5 eq) and tripotassium phosphate (227 mg, 1.07 mmol, 3.0 eq) were added. The mixture was flushed with argon for 5 min before [1,1'-bis(diphenylphosphino)ferrocene]dichloropalladium(II) (13.1 mg, 17.8  $\mu$ mol, 5 mol%) was added and the mixture was heated to 80 °C for 1 h. After cooling to rt H<sub>2</sub>O and EtOAc were added (50 mL each) and the organic layer was separated. The aqueous layer was extracted with EtOAc (3 x 50 mL) and the combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated *in vacuo*. Flash column chromatography (silica, 20 x 2.5 cm, wet load, gradient; *i*Hex/EtOAc = 1:0  $\rightarrow$  100:1  $\rightarrow$  50:1  $\rightarrow$  100:3) yielded the TFA-TBS-protected naphthalene amine nucleoside **111** (140 mg, 0.21 mmol, 45% yield) as a pale yellow oil.

$R_f$  (*i*Hex/EtOAc, 9:1) = 0.6.

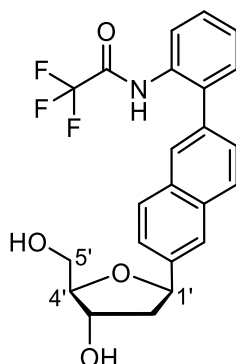
**<sup>1</sup>H NMR** (800 MHz, CDCl<sub>3</sub>):  $\delta$  (ppm) = 8.32 (dd,  $J$  = 8.3, 0.9 Hz, 1H, H<sub>ar</sub>), 8.07 (br s, 1H, NH), 7.95 (d,  $J$  = 8.4 Hz, 1H, H<sub>ar</sub>), 7.84 (d,  $J$  = 8.4 Hz, 1H, H<sub>ar</sub>), 7.60 (dd,  $J$  = 8.4, 1.6 Hz, 1H, H<sub>ar</sub>), 7.58–7.56 (m, 1H, H<sub>ar</sub>), 7.48–7.45 (m, 1H, H<sub>ar</sub>), 7.45–7.42 (m, 2H, H<sub>ar</sub>), 7.42–7.39 (m, 1H, H<sub>ar</sub>), 7.34 (td,  $J$  = 7.5, 1.1 Hz, 1H, H<sub>ar</sub>), 5.35 (dd,  $J$  = 10.5, 5.4 Hz, 1H, 1'-H), 4.53–4.44 (m, 1H, 3'-H), 4.05 (ddd,  $J$  = 5.6, 3.7, 1.9 Hz, 1H, 4'-H), 3.84 (dd,  $J$  = 10.7, 3.7 Hz, 1H, 5'-H), 3.71 (dd,  $J$  = 10.7, 5.7 Hz, 1H, 5'-H), 2.22 (ddd,  $J$  = 12.6, 5.4, 1.4 Hz, 1H, 2'-H), 2.01 (ddd,  $J$  = 12.7, 10.6, 5.3 Hz, 1H, 2'-H), 0.95 (s, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.93 (m, 9H, SiC(CH<sub>3</sub>)<sub>3</sub>), 0.14–0.11 (m, 12H, 2 x Si(CH<sub>3</sub>)<sub>2</sub>).

**<sup>13</sup>C NMR** (201 MHz, <sup>1</sup>H-decoupled, CDCl<sub>3</sub>):  $\delta$  (ppm) = 154.7 (q, <sup>2</sup> $J_{C-F}$  = 37 Hz, CO), 141.3 (C<sub>ar</sub>), 133.9 (C<sub>ar</sub>), 133.2 (C<sub>ar</sub>), 133.1 (C<sub>ar</sub>), 132.9 (C<sub>ar</sub>), 132.4 (C<sub>ar</sub>), 130.8 (HC<sub>ar</sub>), 129.5 (HC<sub>ar</sub>), 129.4 (HC<sub>ar</sub>), 128.9 (HC<sub>ar</sub>), 128.29 (HC<sub>ar</sub>), 126.7 (HC<sub>ar</sub>), 126.5 (HC<sub>ar</sub>), 125.7 (HC<sub>ar</sub>), 121.6 (HC<sub>ar</sub>), 120.6 (HC<sub>ar</sub>), 115.8 (q, <sup>1</sup> $J_{C-F}$  = 289 Hz, CF<sub>3</sub>), 88.5 (4'-C), 80.3 (1'-C), 74.7 (3'-C), 64.1 (5'-C), 44.6 (2'-C), 26.1 (SiC(CH<sub>3</sub>)<sub>3</sub>), 26.0 (SiC(CH<sub>3</sub>)<sub>3</sub>), 18.5 (SiC(CH<sub>3</sub>)<sub>3</sub>), 18.2 (SiC(CH<sub>3</sub>)<sub>3</sub>), -4.45 (SiCH<sub>3</sub>), -4.49 (SiCH<sub>3</sub>), -5.2 (SiCH<sub>3</sub>), -5.3 (SiCH<sub>3</sub>).

**<sup>19</sup>F NMR** (376 MHz, CDCl<sub>3</sub>):  $\delta$  (ppm) = -76.0 (s, 3F).

**HRMS (ESI):** calc. for  $C_{35}H_{47}O_4NF_3Si_2^-$   $[M-H]^-$ : 658.30012; found: 658.29999.

2,2,2-Trifluoro-*N*-(2-(6-((2*R*,4*S*,5*R*)-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)naphthalen-2-yl)phenyl)acetamide (**112**)



$C_{23}H_{20}F_3NO_4$   
MW: 431.40 g/mol

To a stirring solution of the TFA-TBS-protected naphthalene amine nucleoside **111** (170 mg, 258  $\mu$ mol, 1.0 eq) in EtOAc (4.5 mL) in a polypropylene tube HF·pyridine (200  $\mu$ L, 220 mg, 7.7 mmol, 30.0 eq) was added. After 15 h at rt  $Me_3SiOMe$  (5.3 mL, 4.00 g, 38.4 mmol, 150.0 eq) was added dropwise and stirring continued for 30 min before the solution was concentrated *in vacuo*. Flash column chromatography (silica, 20 x 2 cm, wet load, gradient; DCM/MeOH = 1:0  $\rightarrow$  50:1  $\rightarrow$  25:1  $\rightarrow$  9:1) yielded the TFA-protected naphthalene amine nucleoside **112** (77 mg, 178  $\mu$ mol, 69% yield) as a yellow-orange oil.

$R_f$  (DCM/MeOH, 9:1) = 0.6.

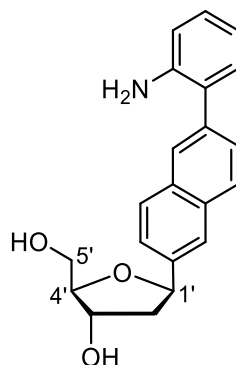
**$^1H$  NMR** (400 MHz,  $CD_3OD$ ):  $\delta$  (ppm) = 7.90–7.77 (m, 4H,  $H_{ar}$ ), 7.56 (dd,  $J$  = 8.5, 1.6 Hz, 1H,  $H_{ar}$ ), 7.53–7.42 (m, 5H,  $H_{ar}$ ), 5.31 (dd,  $J$  = 10.5, 5.4 Hz, 1H, 1'-H), 4.38 (dt,  $J$  = 5.9, 1.8 Hz, 1H, 3'-H), 4.03 (td,  $J$  = 5.1, 2.5 Hz, 1H, 4'-H), 3.74 (dd,  $J$  = 5.1, 1.7 Hz, 2H, 5'-H), 2.29 (ddd,  $J$  = 13.1, 5.5, 1.7 Hz, 1H, 2'-H), 2.06 (ddd,  $J$  = 13.1, 10.5, 5.9 Hz, 1H, 2'-H).

**$^{13}C$  NMR** (101 MHz,  $^1H$ -decoupled,  $CD_3OD$ ):  $\delta$  (ppm) = 157.8 (q,  $^2J_{C-F}$  = 37 Hz, CO), 141.2 ( $C_{ar}$ ), 140.0 ( $C_{ar}$ ), 137.2 ( $C_{ar}$ ), 134.3 ( $C_{ar}$ ), 133.9 ( $C_{ar}$ ), 133.3 ( $C_{ar}$ ), 132.0 ( $HC_{ar}$ ), 129.4 ( $HC_{ar}$ ), 129.30 ( $HC_{ar}$ ), 129.29 ( $HC_{ar}$ ), 129.1 ( $HC_{ar}$ ), 128.6 ( $HC_{ar}$ ), 128.3 ( $HC_{ar}$ ), 128.0 ( $HC_{ar}$ ), 125.9 ( $HC_{ar}$ ), 125.6 ( $HC_{ar}$ ), 117.5 (q,  $^1J_{C-F}$  = 287 Hz,  $CF_3$ ), 89.3 (4'-C), 81.7 (1'-C), 74.5 (3'-C), 64.1 (5'-C), 44.8 (2'-C).

**$^{19}F$  NMR** (376 MHz,  $CD_3OD$ ):  $\delta$  (ppm) = -77.2 (s, 3F).

**HRMS (ESI):** calc. for  $C_{23}H_{19}O_4NF_3^-$   $[M-H]^-$ : 430.12717; found: 430.12735.

(2*R*,3*S*,5*R*)-5-(6-(2-Aminophenyl)naphthalen-2-yl)-2-(hydroxymethyl)tetrahydrofuran-3-ol (**107**)



$C_{21}H_{21}NO_3$   
MW: 335.40 g/mol

To a stirring solution of the TFA-protected naphthalene amine nucleoside **112** (45.0 mg, 104  $\mu$ mol, 1.0 eq) in MeOH/H<sub>2</sub>O (3:1, 0.6 mL) lithium hydroxide hydrate (6.1 mg, 146  $\mu$ mol, 1.4 eq) was added and the solution was warmed to 40 °C for 30 min and then concentrated *in vacuo*. Flash column chromatography (silica, 13 x 1.5 cm, wet load, gradient; DCM/MeOH = 100:1  $\rightarrow$  50:1  $\rightarrow$  20:1  $\rightarrow$  9:1) yielded the naphthalene amine nucleoside **107** (27.6 mg, 82  $\mu$ mol, 79% yield) as a yellow-brown glass.

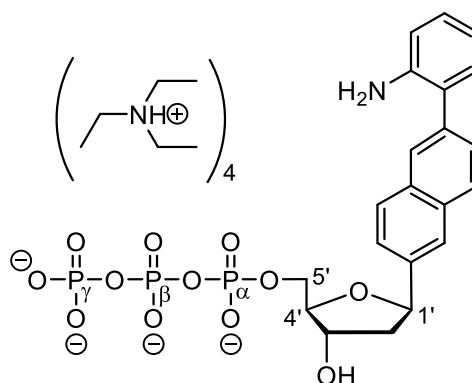
$R_f$  (DCM/MeOH, 9:1) = 0.6.

**<sup>1</sup>H NMR** (400 MHz, CD<sub>3</sub>OD):  $\delta$  (ppm) = 7.92–7.86 (m, 2H, H<sub>ar</sub>), 7.83 (d,  $J$  = 8.7 Hz, 2H, H<sub>ar</sub>), 7.54 (dd,  $J$  = 4.0, 1.8 Hz, 1H, H<sub>ar</sub>), 7.52 (dd,  $J$  = 3.9, 1.7 Hz, 1H, H<sub>ar</sub>), 7.15–7.07 (m, 2H, H<sub>ar</sub>), 6.85 (dd,  $J$  = 8.5, 1.2 Hz, 1H, H<sub>ar</sub>), 6.79 (td,  $J$  = 7.5, 1.2 Hz, 1H, H<sub>ar</sub>), 5.30 (dd,  $J$  = 10.4, 5.4 Hz, 1H, 1'-H), 4.37 (dt,  $J$  = 5.9, 1.8 Hz, 1H, 3'-H), 4.02 (td,  $J$  = 5.1, 2.5 Hz, 1H, 4'-H), 3.86–3.55 (m, 2H, 5'-H), 2.27 (ddd,  $J$  = 13.1, 5.3, 1.7 Hz, 1H, 2'-H), 2.05 (ddd,  $J$  = 13.1, 10.5, 5.9 Hz, 1H, 2'-H).

**<sup>13</sup>C NMR** (101 MHz, <sup>1</sup>H-decoupled, CD<sub>3</sub>OD):  $\delta$  (ppm) = 145.4 (*ipso*-C), 140.8 (*ipso*-C), 138.7 (*ipso*-C), 134.7 (*ipso*-C), 133.7 (*ipso*-C), 131.5 (HC<sub>ar</sub>), 129.5 (HC<sub>ar</sub>), 129.4 (HC<sub>ar</sub>), 129.2 (HC<sub>ar</sub>), 129.0 (*ipso*-C), 128.6 (HC<sub>ar</sub>), 128.4 (HC<sub>ar</sub>), 125.72 (HC<sub>ar</sub>), 125.66 (HC<sub>ar</sub>), 119.6 (HC<sub>ar</sub>), 117.2 (HC<sub>ar</sub>), 89.3 (4'-C), 81.7 (1'-C), 74.5 (3'-C), 64.1 (5'-C), 44.9 (2'-C).

**HRMS (ESI)**: calc. for C<sub>21</sub>H<sub>22</sub>O<sub>3</sub>N<sup>+</sup> [M+H]<sup>+</sup>: 336.15942; found: 336.15941.

((2*R*,3*S*,5*R*)-5-(6-(2-Aminophenyl)naphthalen-2-yl)-3-hydroxytetrahydrofuran-2-yl)methyl triphosphate  
(**113**)



$C_{21}H_{24}NO_{12}P_3$  (hypothetic free acid)  
MW: 575.34 g/mol

To the excessively dried naphthalene amine nucleoside **107** (27.6 mg, 82.3  $\mu$ mol, 1.0 eq) under argon was added the triphosphate reagent solution (0.87 mL, 165.3  $\mu$ mol, 2.0 eq, prepared according to general procedure **1**, p. 100) at 0 °C and the reaction mixture was allowed to warm to rt under stirring. After 2.5 h TLC (DCM/MeOH = 5:1) indicated the presence of unreacted naphthalene amine nucleoside, so more triphosphate reagent solution (0.20 mL, 38.0  $\mu$ mol, 0.47 eq) was added and stirring continued for 30 min. The reaction mixture was oxidized with an iodine solution (20 mM  $I_2$  in Py/ $H_2O$  = 9:1) until a slight brown color remained for 15 min (approx. 1.0 mL), followed by dd $H_2O$  (2.7 mL) and stirred for 1 h. Then an aqueous solution of NaCl (3 M, 1.6 mL) was added, and the reaction mixture was transferred to a 50 mL centrifugation tube, vortexed vigorously and mixed with EtOH (absolute, 15 mL). Precipitation was achieved by cooling to –80 °C for 30 min, followed by centrifugation (5 min, 3200 x g). The supernatant was discarded and the pellet was redissolved in buffer A (1.5 mL) and lyophilized. The residue was purified twice by RP-HPLC purification (0–35% B in 45 min). This afforded the naphthalene amine triphosphate **113** as a slightly yellow tetrakis-triethylammonium salt (16.0 mg, 14.1  $\mu$ mol, 17%) after freeze-drying.

**$^1H$  NMR** (400 MHz,  $D_2O$ ):  $\delta$  (ppm) = 7.91–7.84 (m, 2H,  $H_{ar}$ ), 7.82 (d,  $J$  = 8.6 Hz, 1H,  $H_{ar}$ ), 7.70 (s, 1H,  $H_{ar}$ ), 7.55 (d,  $J$  = 8.6 Hz, 1H,  $H_{ar}$ ), 7.38 (d,  $J$  = 8.5 Hz, 1H,  $H_{ar}$ ), 7.24 (t,  $J$  = 7.7 Hz, 1H,  $H_{ar}$ ), 7.10 (d,  $J$  = 7.5 Hz, 1H,  $H_{ar}$ ), 7.00 (d,  $J$  = 8.0 Hz, 1H,  $H_{ar}$ ), 6.95 (t,  $J$  = 7.5 Hz, 1H,  $H_{ar}$ ), 5.27 (dd,  $J$  = 10.3, 5.7, 1H, 1'-H), 4.58–4.52 (m, 1H, 3'-H), 4.23–4.17 (m, 1H, 4'-H), 4.16–4.05 (m, 2H, 5'-H), 2.30–2.11 (m, 2H, 2'-H).

**$^{31}P$  NMR** (162 MHz,  $D_2O$ ):  $\delta$  (ppm) = –11.02 (d,  $^2J_{P-P}$  = 19.8 Hz,  $\gamma$ -P), –11.33 (dt,  $^2J_{P-P}$  = 20.1 Hz,  $^3J_{P-H}$  = 5.9 Hz,  $\alpha$ -P), –23.46 (t,  $^2J_{P-P}$  = 20.0 Hz,  $\beta$ -P).

**HRMS (ESI)**: calc. for  $C_{21}H_{23}O_{12}NP_3^-$  [ $M-H$ ] $^-$ : 574.04386; found: 574.04393.

**UV-Vis** ( $H_2O$ ):  $\lambda_{Abs}$  (nm) = 230 (base peak), 279.

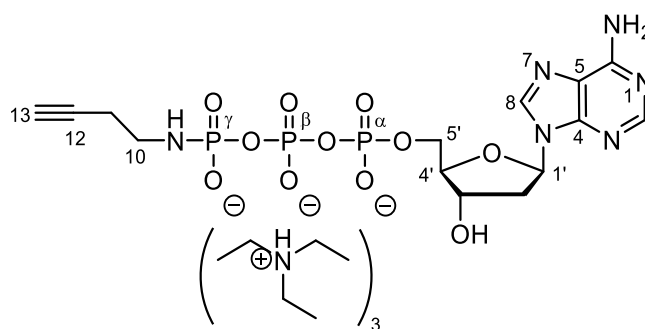
**Extinction coefficient** ( $H_2O$ ):  $\epsilon$  (279 nm) = 9000  $M^{-1}cm^{-1}$ .

### 5.2.9 Synthesis of $\gamma$ -Alkyne Labeled Nucleotides

**General Procedure 2:** In a 1.5 mL reaction tube 300  $\mu$ L of the nucleoside triphosphate solution (100 mM, 30  $\mu$ mol, 1.0 eq) were mixed with 300  $\mu$ L EDC·HCl solution (500 mM in ddH<sub>2</sub>O, pH = 7.5, 150  $\mu$ mol, 5.0 eq) and incubated at 25 °C, 1200 rpm (in a thermomixer from EPPENDORF) for 7 min. Then, 600  $\mu$ L of 1-aminobut-3-yne solution (55 mM in DMF, 33  $\mu$ mol, 1.1 eq) were added and incubated for 4.5 h at 25 °C, 1200 rpm. DMF was removed by extraction with CHCl<sub>3</sub> (3 x 600  $\mu$ L) and the aqueous layer (aprox. 600  $\mu$ L) was transferred to a 15 mL reaction tube. Precipitation was achieved by addition of 180  $\mu$ L NaCl (3 M in ddH<sub>2</sub>O) and 4 mL of abs. EtOH. After vortexing, the mixture was cooled to –80 °C for 1 h (alternatively –20 °C overnight). After centrifugation (4000 x g, 10 min) the supernatant was removed, the colorless solid was redissolved in H<sub>2</sub>O (500  $\mu$ L) and then dried by lyophilization. Purification by preparative RP-HPLC yielded  $\gamma$ -alkyne labeled nucleotides with >95% purity (according to analytical RP-HPLC at 260 nm detection, 0–20% B in 45 min).

**General Procedure 3:** All reactions were performed with magnetic stirring. EDC·HCl (5.0 eq) was dissolved in ddH<sub>2</sub>O, the nucleoside triphosphate (1.0 eq) was added and the pH was adjusted to 7.5 with aqueous NaOH (0.1 M) if necessary. After stirring at rt for 5 min 1-aminobut-3-yne (55 mM in DMF, 1.1 eq) was added. After complete consumption of the starting material (reactions were monitored by RP-HPLC at 260 nm detection, 0–20% B in 45 min), the solvent was removed *in vacuo* and the colorless crude product was purified by preparative RP-HPLC to yield the  $\gamma$ -alkyne labeled nucleotides.

$\gamma$ -N-(But-3-yn-1-ylamido)-2'-deoxyadenosine-5'-triphosphate **118**



C<sub>14</sub>H<sub>21</sub>N<sub>6</sub>O<sub>11</sub>P<sub>3</sub> (hypothetic free acid)  
MW: 542.27 g/mol

The synthesis was performed with a 100 mM dATP solution (300  $\mu$ L, 30.0  $\mu$ mol) as described in the general procedure **2** (5.2.9, p. 130). RP-HPLC purification (0–15% B in 45 min) afforded the triethylammonium salt of  $\gamma$ -alkyne labeled dATP **118** (19.9 mg, 23.5  $\mu$ mol, 78%) as a colorless solid after lyophilization.

<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O):  $\delta$  (ppm) = 8.56 (s, 1H, 2-H), 8.31 (s, 1H, 8-H), 6.55 (t, <sup>3</sup>J = 4.0 Hz, 1H, 1'-H), 4.84–4.82 (m, 1H, 3'-H), 4.33–4.25 (m, 1H, 4'-H), 4.33–4.10 (m, 2H, 5'-H), 3.10–2.99 (m, 2H, 10-H),



2.89–2.83 (m, 1H, 2'-H), 2.67–2.60 (m, 1H, 2'-H), 2.36 (td,  $^3J = 6.4$  Hz,  $^4J = 2.6$  Hz, 2H, 11-H), 2.31 (t,  $^4J = 2.4$  Hz, 1H, 13-H).

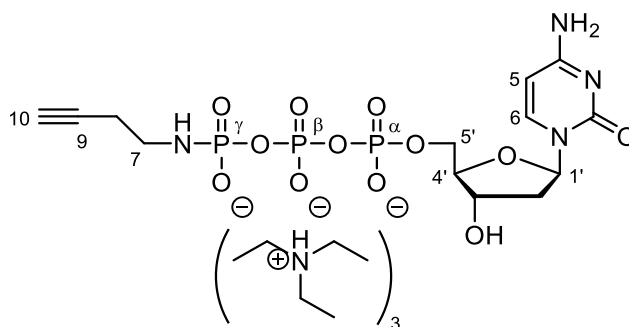
$^{13}\text{C}$  NMR (102 MHz,  $^1\text{H}$ -decoupled,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = 155.3 (6-C), 152.5 (2-C), 148.7 (4-C), 140.0 (8-C), 118.5 (5-C), 85.7 (d,  $^3J_{\text{P-C}} = 9.2$  Hz, 4'-C), 83.6 (1'-C), 83.2 (12-C), 72.3 (3'-C), 71.0 (13-C), 65.4 (d,  $^2J_{\text{P-C}} = 7.1$  Hz, 5'-C), 40.4 (10-C), 39.0 (2'-C), 20.8 (d,  $^3J_{\text{P-C}} = 8.2$  Hz, 11-C).

$^{31}\text{P}$  NMR (162 MHz,  $^1\text{H}$ -decoupled,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = -2.00 (d,  $^2J = 19.4$  Hz,  $\gamma\text{-P}$ ), -11.58 (d,  $^2J = 19.6$  Hz,  $\alpha\text{-P}$ ), -23.15 (t,  $^2J = 21.6$  Hz,  $\beta\text{-P}$ ).

IR (ATR):  $\tilde{\nu}$  ( $\text{cm}^{-1}$ ) = 3177, 2982, 2884, 2610, 2471, 1650, 1597, 1570, 1473, 1397, 1330, 1294, 1212, 1083, 991, 901, 836, 799, 723.

**HRESIMS**: calculated for  $\text{C}_{14}\text{H}_{20}\text{N}_6\text{O}_{11}\text{P}_3^-$  [ $\text{M-H}$ ] $^-$ : 541.0408, observed: 541.0399.

$\gamma$ -N-(But-3-yn-1-ylamido)-2'-deoxycytidine-5'-triphosphate **119**



$\text{C}_{13}\text{H}_{21}\text{N}_4\text{O}_{12}\text{P}_3$  (hypothetic free acid)  
MW: 518.25 g/mol

The synthesis was performed with a 100 mM dCTP solution (300  $\mu\text{L}$ , 30.0  $\mu\text{mol}$ ) as described in the general procedure **2** (5.2.9, p. 130). RP-HPLC purification (0–10% B in 45 min) afforded the tri-triethylammonium salt of  $\gamma$ -alkyne labeled dCTP **119** (18.9 mg, 23.0  $\mu\text{mol}$ , 76%) as a colorless solid after lyophilization.

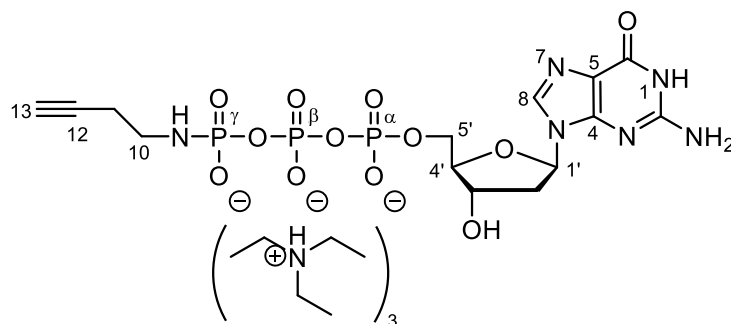
$^1\text{H}$  NMR (400 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  = 8.04 (d,  $^3J = 7.53$  Hz, 1H, 6-H), 6.37 (t,  $^3J = 6.92$  Hz, 1H, 1'-H), 6.19 (d,  $^3J = 7.57$  Hz, 1H, 5-H), 4.67–4.64 (m, 1H, 3'-H), 4.25–4.24 (m, 3H, 4'-H, 5'-H<sub>2</sub>), 3.12–3.06 (m, 2H, 7-H), 2.69–2.65 (m, 1H, 2'-H), 2.47–2.46 (m, 1H, 2'-H), 2.45–2.40 (m, 2H, 8-H), 2.38–2.36 (m, 1H, 10-H).

$^{13}\text{C}$  NMR (102 MHz,  $^1\text{H}$ -decoupled,  $\text{D}_2\text{O}$ ):  $\delta$  = 166.0 (4-C), 157.3 (2-C), 141.7 (6-C), 96.5 (1'-C), 85.8 (5-C), 85.4 (d,  $^3J_{\text{P-C}} = 9.6$  Hz, 4'-C), 83.3 (9-C), 70.4 (3'-C), 70.1 (10-C), 65.0 (d,  $^2J_{\text{P-C}} = 5.6$  Hz, 5'-C), 40.4 (7-C), 39.4 (2'-C), 20.7 ( $^3J_{\text{P-C}} = 8.7$  Hz, 8-C).

$^{31}\text{P}$  NMR (162 MHz,  $^1\text{H}$ -decoupled,  $\text{D}_2\text{O}$ ):  $\delta$  = -1.97 (d,  $^2J = 20.7$  Hz,  $\gamma\text{-P}$ ), -11.63 (d,  $^2J = 19.5$  Hz,  $\alpha\text{-P}$ ), -23.10 (t,  $^2J = 20.6$  Hz,  $\beta\text{-P}$ ).

IR (ATR):  $\tilde{\nu}$  ( $\text{cm}^{-1}$ ) = 3276, 2973, 2937, 2877, 2437, 1692, 1647, 1524, 1488, 1385, 1223, 1057, 995, 902, 840, 809, 787, 767.

**HRESIMS**: calculated for  $\text{C}_{13}\text{H}_{20}\text{N}_4\text{O}_{12}\text{P}_3^-$  [ $\text{M-H}$ ] $^-$ : 517.0296, observed: 517.0296.

$\gamma$ -*N*-(But-3-yn-1-ylamido)-2'-deoxyguanosine-5'-triphosphate **120**

$\text{C}_{14}\text{H}_{21}\text{N}_6\text{O}_{12}\text{P}_3$  (hypothetic free acid)  
MW: 558.27 g/mol

The synthesis was performed with a 100 mM dGTP solution (300  $\mu\text{L}$ , 30.0  $\mu\text{mol}$ ) as described in the general procedure **2** (5.2.9, p. 130) except that after EDC·HCl incubation 200  $\mu\text{L}$  of extra  $\text{H}_2\text{O}$  were added to prevent precipitation upon addition of the DMF solution. RP-HPLC purification (0–15% B in 45 min) afforded the tris-triethylammonium salt of  $\gamma$ -alkyne labeled dGTP **120** (20.3 mg, 23.6  $\mu\text{mol}$ , 78%) as a colorless solid after lyophilization.

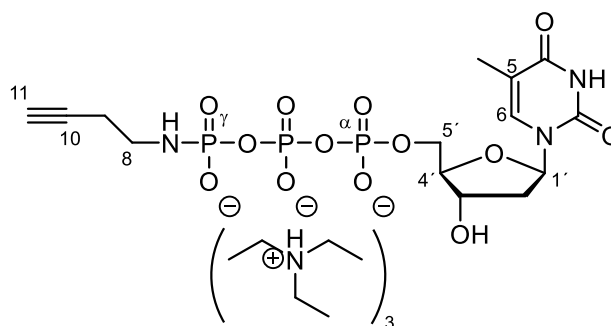
**$^1\text{H}$  NMR** (400 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = 8.11 (s, 1H, 8-H), 6.33 (t,  $^3J = 4.6$  Hz, 1H, 1'-H), 4.81–4.78 (m, 1H, 3'-H), 4.28–4.25 (m, 1H, 4'-H), 4.20 (t,  $^3J = 3.6$  Hz, 2H, 5'-H), 3.10–3.01 (m, 2H, 10-H), 2.89–2.83 (m, 1H, 2'-H), 2.53–2.48 (m, 1H, 2'-H), 2.37–2.34 (m, 2H, 11-H), 2.31–2.29 (m, 1H, 13-H).

**$^{13}\text{C}$  NMR** (150.6 MHz,  $^1\text{H}$ -decoupled,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = 158.9 (6-C), 153.8 (2-C), 151.4 (4-C), 137.8 (8-C), 116.3 (5-C), 85.6 (d,  $^3J_{\text{P-C}} = 9.2$  Hz, 4'-C), 83.6 (1'-C), 83.3 (12-C), 71.2 (3'-C), 70.0 (13-C), 65.3 (d,  $^2J_{\text{P-C}} = 5.7$  Hz, 5'-C), 40.4 (10-C), 38.2 (2'-C), 20.7 (d,  $^3J_{\text{P-C}} = 8.4$  Hz, 11-C).

**$^{31}\text{P}$  NMR** (162 MHz,  $^1\text{H}$ -decoupled,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = -1.94 (d,  $^2J = 21.6$  Hz,  $\gamma$ -P), -11.51 (d,  $^2J = 19.8$  Hz,  $\alpha$ -P), -23.08 (t,  $^2J = 20.3$  Hz,  $\beta$ -P).

**IR** (ATR):  $\tilde{\nu}$  ( $\text{cm}^{-1}$ ) = 3212, 2986, 2945, 2694, 2497, 1678, 1635, 1602, 1568, 1531, 1478, 1454, 1397, 1358, 1320, 1225, 1085, 1060, 996, 911, 837, 783, 734, 677.

**HRESIMS**: calculated for  $\text{C}_{14}\text{H}_{20}\text{N}_6\text{O}_{13}\text{P}_3^-$  [ $\text{M-H}$ ] $^-$ : 557.0358, observed: 557.0346.

$\gamma$ -*N*-(But-3-yn-1-ylamido)-2'-deoxythymidine-5'-triphosphate **121**C<sub>14</sub>H<sub>22</sub>N<sub>3</sub>O<sub>13</sub>P<sub>3</sub> (hypothetic free acid)

MW: 533.26 g/mol

The synthesis was performed with dTTP sodium salt (13.8 mg, 25.0  $\mu$ mol, 1.0 eq), EDC·HCl (24.0 mg, 125.0  $\mu$ mol, 5.0 eq) and 1-aminobut-3-yne (1.9 mg, 28.0  $\mu$ mol, 1.1 eq) in ddH<sub>2</sub>O (0.5 mL) within 4.5 h as described in the general procedure **3** (5.2.9, p. 130). RP-HPLC purification (0–15% B in 45 min) afforded the tris-triethylammonium salt of  $\gamma$ -alkyne labeled dTTP **121** (17.5 mg, 21.0  $\mu$ mol, 84%) as a colorless solid after lyophilization.

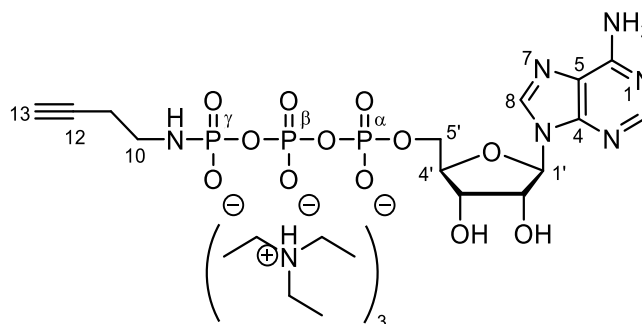
**<sup>1</sup>H NMR** (400 MHz, D<sub>2</sub>O):  $\delta$  (ppm) = 7.80 (d,  $^3J$  = 7.7 Hz, 1H, 6-H), 6.37 (t,  $^3J$  = 6.9 Hz, 1H, 1'-H), 4.70–4.66 (m, 1H, 3'-H), 4.26–4.18 (m, 3H, 4'-H, 5'-H), 3.11–3.05 (m, 2H, 8-H), 2.44–2.33 (m, 5H, 2'-H, 9-H, 11-H), 1.96 (s, 3H, 7-H).

**<sup>13</sup>C NMR** (102 MHz, <sup>1</sup>H-decoupled, D<sub>2</sub>O):  $\delta$  (ppm) = 166.5 (4-C), 151.7 (2-C), 137.3 (6-C), 111.7 (5-C), 85.4 (d,  $^3J_{P-C}$  = 9.1 Hz, 4'-C), 84.8 (1'-C), 83.4 (10-C), 70.7 (3'-C), 70.1 (11-C), 65.3 (d,  $^2J_{P-C}$  = 5.7 Hz, 5'-C), 40.4 (8-C), 38.5 (2'-C), 20.8 (d,  $^3J_{P-C}$  = 8.6 Hz, 9-C), 11.6 (7-C).

**<sup>31</sup>P NMR** (162 MHz, D<sub>2</sub>O):  $\delta$  (ppm) = –1.94 (d,  $^2J_{P-P}$  = 20.6 Hz,  $\gamma$ -P), –11.85 (dd,  $^2J_{P-P}$  = 19.9 Hz,  $^3J_{H-P}$  = 3.8 Hz,  $\alpha$ -P), –23.15 (t,  $^2J_{P-P}$  = 20.2 Hz,  $\beta$ -P).

**IR** (ATR):  $\tilde{\nu}$  (cm<sup>–1</sup>) = 3250, 2914, 2847, 2692, 2493, 1660, 1463, 1399, 1223, 1060, 991, 906, 814, 764, 717.

**HRESIMS**: calculated for C<sub>14</sub>H<sub>21</sub>N<sub>3</sub>O<sub>13</sub>P<sub>3</sub><sup>–</sup> [M–H]<sup>–</sup>: 532.0293, observed: 532.0291.

 $\gamma$ -*N*-(But-3-yn-1-ylamido)adenosine-5'-triphosphate **122**C<sub>14</sub>H<sub>21</sub>N<sub>6</sub>O<sub>12</sub>P<sub>3</sub> (hypothetic free acid)

MW: 558.27 g/mol

The synthesis was performed with ATP disodium salt (27.6 mg, 50.0  $\mu\text{mol}$ , 1.0 eq), EDC·HCl (48.0 mg, 250.0  $\mu\text{mol}$ , 1.0 eq) and 1-aminobut-3-yne (3.8 mg, 55.0  $\mu\text{mol}$ , 1.1 eq) in ddH<sub>2</sub>O (1.0 mL) within 3 h as described in the general procedure **3** (5.2.9, p. 130). RP-HPLC purification (0–15% B in 45 min) afforded the tris-triethylammonium salt of  $\gamma$ -alkyne labeled ATP **122** (36.7 mg, 43.0  $\mu\text{mol}$ , 85%) as a colorless solid after lyophilization.

**<sup>1</sup>H NMR** (400 MHz, D<sub>2</sub>O):  $\delta$  (ppm) = 8.60 (s, 1H, 2-H), 8.31 (s, 1H, 8-H), 6.18 (d,  $^3J = 6.0$  Hz, 1H, 1'-H), 4.84–4.80 (m, 1H, 2'-H), 4.62–4.60 (m, 1H, 3'-H), 4.44–4.43 (m, 1H, 4'-H), 4.30–4.26 (m, 2H, 5'-H), 3.07–3.03 (m, 2H, 10-H), 2.38–2.26 (m, 3H, 11-H, 13-H).

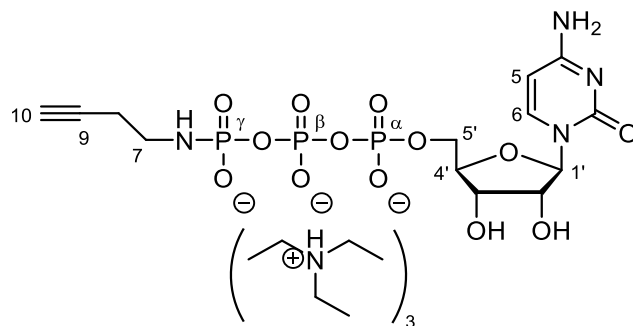
**<sup>13</sup>C NMR** (102 MHz, <sup>1</sup>H-decoupled, D<sub>2</sub>O):  $\delta$  (ppm) = 155.4 (6-C), 152.7 (2-C), 149.3 (4-C), 140.0 (8-C), 118.4 (5-C), 86.6 (1'-C), 84.1 (d,  $^3J_{\text{P-C}} = 9.3$  Hz, 4'-C), 83.3 (12-C), 74.2 (2'-C), 70.4 (3'-C), 70.3 (13-C), 65.2 (d,  $^2J_{\text{P-C}} = 6.1$  Hz, 5'-C), 40.4 (10-C), 20.8 (d,  $^3J_{\text{P-C}} = 8.5$  Hz, 11-C).

**<sup>31</sup>P NMR** (162 MHz, <sup>1</sup>H-decoupled, D<sub>2</sub>O):  $\delta$  (ppm) = -1.94 (d,  $^2J = 20.7$  Hz,  $\gamma$ -P), -11.58 (d,  $^2J = 19.6$  Hz,  $\alpha$ -P), -23.03 (t,  $^2J = 20.2$  Hz,  $\beta$ -P).

**IR** (ATR):  $\tilde{\nu}$  (cm<sup>-1</sup>) = 3189, 2982, 2680, 2489, 1645, 1600, 1571, 1475, 1399, 1331, 1297, 1215, 1060, 990, 897, 799.

**HRESIMS**: calculated for C<sub>14</sub>H<sub>20</sub>N<sub>6</sub>O<sub>12</sub>P<sub>3</sub><sup>-</sup> [M-H]<sup>-</sup>: 557.0358, observed: 557.0355.

#### $\gamma$ -N-(But-3-yn-1-ylamido)cytidine-5'-triphosphate **123**



C<sub>13</sub>H<sub>21</sub>N<sub>4</sub>O<sub>13</sub>P<sub>3</sub> (hypothetic free acid)  
MW: 534.25 g/mol

The synthesis was performed with a 100 mM CTP solution (300  $\mu\text{L}$ , 30.0  $\mu\text{mol}$ ) as described in the general procedure **2** (5.2.9, p. 130). RP-HPLC purification (0–10% B in 45 min) afforded the tris-triethylammonium salt of  $\gamma$ -alkyne labeled CTP **123** (19.2 mg, 23.0  $\mu\text{mol}$ , 76%) as a colorless solid after lyophilization.

**<sup>1</sup>H NMR** (400 MHz, D<sub>2</sub>O):  $\delta$  (ppm) = 8.08 (d,  $^3J = 8.0$  Hz, 1H, 6-H), 6.22 (d,  $^3J = 8.0$  Hz, 1H, 5-H), 6.03 (d,  $^3J = 4.4$  Hz, 1H, 1'-H), 4.43–4.41 (m, 1H, 3'-H), 4.37–4.34 (m, 1H, 2'-H), 4.32–4.25 (m, 3H, 4'-H, 5'-H), 3.08 (dt,  $^2J_{\text{P-H}} = 10.4$  Hz,  $^3J = 6.8$  Hz, 2H, 7-H), 2.41 (td,  $^3J = 6.8$  Hz,  $^4J = 2.6$  Hz, 2H, 8-H), 2.37 (t,  $^4J = 2.6$  Hz, 1H, 10-H).

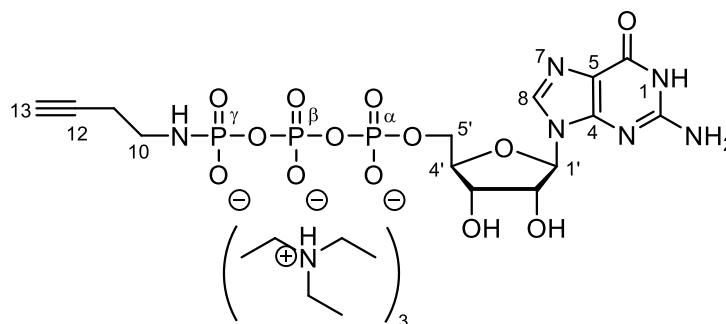
**$^{13}\text{C}$  NMR** (102 MHz,  $^1\text{H}$ -decoupled,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = 165.6 (4-C), 157.2 (2-C), 141.6 (6-C), 96.6 (5-C), 89.0 (1'-C), 83.3 (9-C), 82.7 (d,  $^3J_{\text{P-C}} = 9.6$  Hz, 4'-C), 74.2 (2'-C), 70.0 (10-C), 69.0 (3'-C), 64.4 (d,  $^2J_{\text{P-C}} = 5.6$  Hz, 5'-C), 40.4 (7-C), 20.8 (d,  $^3J_{\text{P-C}} = 8.7$  Hz, 8-C).

**$^{31}\text{P}$  NMR** (162 MHz,  $^1\text{H}$ -decoupled,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = -1.98 (d,  $^2J = 20.7$  Hz,  $\gamma\text{-P}$ ), -11.63 (d,  $^2J = 19.5$  Hz,  $\alpha\text{-P}$ ), -23.09 (t,  $^2J = 20.5$  Hz,  $\beta\text{-P}$ ).

**IR** (ATR):  $\tilde{\nu}$  ( $\text{cm}^{-1}$ ) = 3248, 2994, 2502, 1645, 1489, 1398, 1285, 1220, 1109, 1061, 1011, 903, 902, 787.

**HRESIMS**: calculated for  $\text{C}_{13}\text{H}_{20}\text{N}_4\text{O}_{13}\text{P}_3^-$   $[\text{M-H}]^-$ : 533.0245, observed: 533.0233.

$\gamma$ -*N*-(But-3-yn-1-ylamido)guanosine-5'-triphosphate **124**



$\text{C}_{14}\text{H}_{21}\text{N}_6\text{O}_{13}\text{P}_3$  (hypothetic free acid)  
MW: 574.27 g/mol

The synthesis was performed with a 100 mM GTP solution (300  $\mu\text{L}$ , 30.0  $\mu\text{mol}$ ) as described in the general procedure **2** (5.2.9, p. 130). RP-HPLC purification (0–15% B in 45 min) afforded the tri-triethylammonium salt of  $\gamma$ -alkyne labeled GTP **124** (22.8 mg, 26.0  $\mu\text{mol}$ , 86%) as a colorless solid after lyophilization.

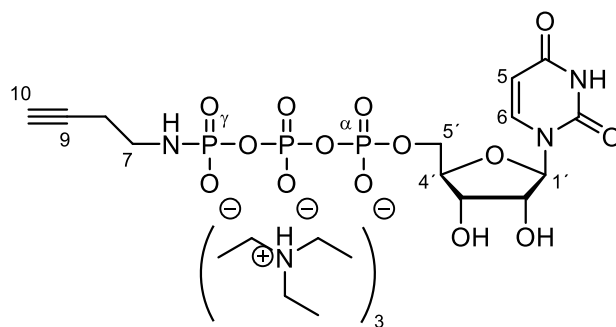
**$^1\text{H}$  NMR** (400 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = 8.15 (s, 1H, 8-H), 5.94 (d,  $^3J = 6.4$  Hz, 1H, 1'-H), 4.87–4.82 (m, 1H, 2'-H), 4.61–4.57 (m, 1H, 3'-H), 4.39–4.35 (m, 1H, 4'-H), 4.30–4.20 (m, 2H, 5'-H), 3.09–3.01 (m, 2H, 10-H), 2.40–2.34 (m, 2H, 11-H), 2.33–2.31 (m, 1H, 13-H).

**$^{13}\text{C}$  NMR** (102 MHz,  $^1\text{H}$ -decoupled,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = 159.0 (6-C), 154.0 (2-C), 151.8 (4-C), 137.7 (8-C), 116.3 (5-C), 86.7 (1'-C), 83.9 (d,  $^3J_{\text{P-C}} = 9.2$  Hz, 4'-C), 82.9 (12-C), 73.4 (2'-C), 70.3 (3'-C and 13-C), 65.2 (d,  $^2J_{\text{P-C}} = 5.6$  Hz, 5'-C), 40.3 (10-C), 20.7 (d,  $^2J_{\text{P-C}} = 8.5$  Hz, 11-C).

**$^{31}\text{P}$  NMR** (162 MHz,  $^1\text{H}$ -decoupled,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = -1.97 (d,  $^2J = 20.6$  Hz,  $\gamma\text{-P}$ ), -11.64 (d,  $^2J = 19.6$  Hz,  $\alpha\text{-P}$ ), -23.12 (t,  $^2J = 20.2$  Hz,  $\beta\text{-P}$ ).

**IR** (ATR):  $\tilde{\nu}$  ( $\text{cm}^{-1}$ ) = 3301, 2986, 2948, 2632, 2491, 1678, 1567, 1532, 1477, 1454, 1398, 1223, 1116, 1061, 1011, 912, 837, 810, 783.

**HRESIMS**: calculated for  $\text{C}_{14}\text{H}_{20}\text{N}_6\text{O}_{13}\text{P}_3^-$   $[\text{M-H}]^-$ : 573.0307, observed: 573.0307.

$\gamma$ -*N*-(But-3-yn-1-ylamido)uridine-5'-triphosphate **125**

$\text{C}_{13}\text{H}_{20}\text{N}_3\text{O}_{14}\text{P}_3$  (hypothetic free acid)  
MW: 535.23 g/mol

The synthesis was performed with a 100 mM UTP solution (300  $\mu\text{L}$ , 30.0  $\mu\text{mol}$ ) as described in the general procedure **2** (5.2.9, p. 130). RP-HPLC purification (0–15% B in 45 min) afforded the tris-triethylammonium salt of  $\gamma$ -alkyne labeled UTP **125** (17.6 mg, 21.0  $\mu\text{mol}$ , 70%) as a colorless solid after lyophilization.

**$^1\text{H}$  NMR** (400 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = 8.02 (d,  $^3J = 7.7$  Hz, 1H, 6-H), 6.02 (d,  $^3J = 7.6$  Hz, 1H, 5-H), 6.02–5.99 (m, 1H, 1'-H), 4.46–4.41 (m, 2H, 3'-H, 2'-H), 4.33–4.29 (m, 1H, 4'-H), 4.28–4.25 (m, 2H, 5'-H), 3.12–3.05 (m, 2H, 7-H), 2.44–2.40 (m, 2H, 8-H), 2.38–2.37 (m, 1H, 10-H).

**$^{13}\text{C}$  NMR** (102 MHz,  $^1\text{H}$ -decoupled,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = 166.1 (4-C), 151.8 (2-C), 141.7 (6-C), 102.7 (5-C), 88.1 (1'-C), 83.3 (d,  $^3J_{\text{P-C}} = 9.8$  Hz, 4'-C), 83.2 (9-C), 73.7 (2'-C), 70.1 (10-C), 69.6 (3'-C), 64.8 (d,  $^2J_{\text{P-C}} = 5.6$  Hz, 5'-C), 40.4 (7-C), 20.8 (d,  $^3J_{\text{P-C}} = 8.2$  Hz, 8-C).

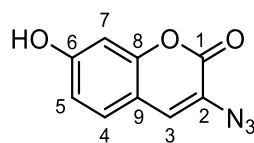
**$^{31}\text{P}$  NMR** (162 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = –1.95 (d,  $^2J_{\text{P-P}} = 21.0$  Hz,  $\gamma$ -P), –11.65 (dd,  $^2J_{\text{P-P}} = 19.4$  Hz,  $^3J_{\text{H-P}} = 3.8$  Hz,  $\alpha$ -P), –23.10 (t,  $^2J_{\text{P-P}} = 21.0$  Hz,  $\beta$ -P).

**IR** (ATR):  $\tilde{\nu}$  ( $\text{cm}^{-1}$ ) = 3259, 2988, 2690, 2504, 1680, 1463, 1390, 1220, 1009, 902, 813, 764.

**HRESIMS**: calculated for  $\text{C}_{13}\text{H}_{20}\text{N}_3\text{O}_{14}\text{P}_3^-$   $[\text{M-H}]^-$ : 534.0085, observed: 534.0074.

### 5.2.10 Synthesis of Fluorophore Azides

#### 3-Azido-7-hydroxy-2H-chromen-2-one (b)



MW: 203.15 g/mol

The synthesis was performed according to the procedure of Sivakumar *et al.*<sup>[216]</sup>

**<sup>1</sup>H NMR** (400 MHz, d<sub>6</sub>-DMSO):  $\delta$  (ppm) = 10.52 (s, 1H, OH), 7.60 (s, 1H, 3-H), 7.48 (d, <sup>3</sup>*J* = 8.5 Hz, 1H, 4-H), 6.81 (dd, <sup>3</sup>*J* = 8.5 Hz, <sup>4</sup>*J* = 2.3 Hz, 1H, 5-H), 6.76 (d, <sup>4</sup>*J* = 2.3 Hz, 1H, 7-H).

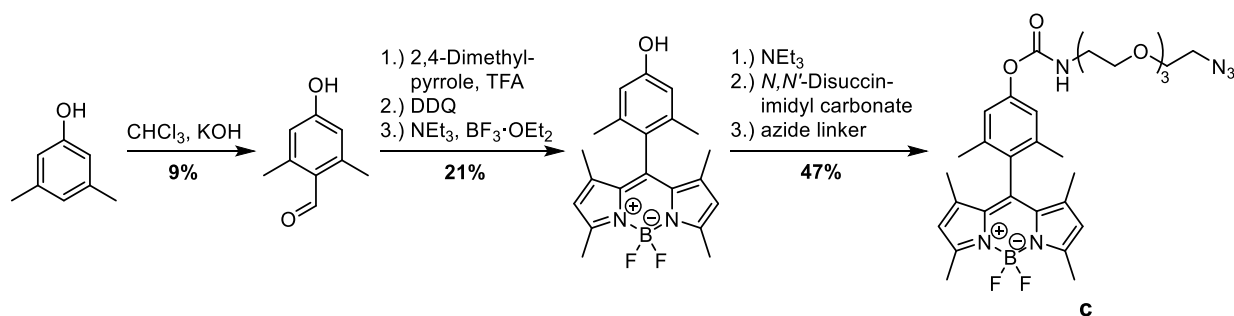
**<sup>13</sup>C NMR** (102 MHz, <sup>1</sup>H-decoupled, d<sub>6</sub>-DMSO):  $\delta$  (ppm) = 160.2 (1-C), 157.3 (6-C), 152.7 (8-C), 129.1 (3-C), 127.8 (4-C), 121.1 (2-C), 113.8 (9-C), 111.3 (5-C), 102.0 (7-C).

**IR** (ATR):  $\tilde{\nu}$  (cm<sup>-1</sup>) = 3204, 2920, 2100, 1739, 1608, 1544, 1505, 1447, 1307, 1243, 1175, 1108, 993, 940, 837, 761, 661.

**HREIMS**: calculated for C<sub>9</sub>H<sub>5</sub>N<sub>3</sub>O<sub>3</sub><sup>+</sup> [M]<sup>+</sup>: 203.0331, observed: 203.0321.

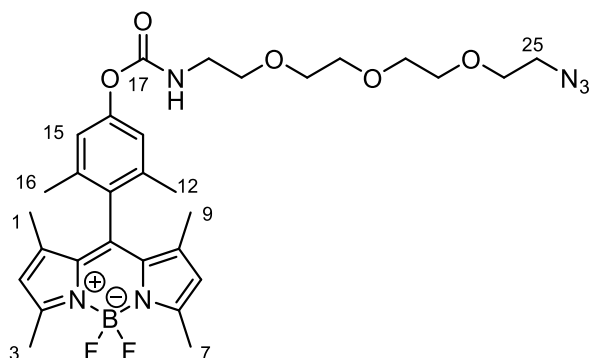
**UV-Vis** (DMSO):  $\lambda_{\text{Abs}}$  (nm) = 345.

#### Synthesis Overview of BODIPY Azide c



**Figure 5-1.** Synthesis of BODIPY azide **c**. Note that BODIPY phenol was generously provided by B. Hackner.

10-(4-(2-(2-(2-(2-Azidoethoxy)ethoxy)ethoxy)ethylcarbamoyloxy)-2,6-di-methylphenyl)-5,5-difluoro-1,3,7,9-tetramethyl-5*H*-dipyrrolo[1,2-*c*:1',2'-*f*][1,3,2]diazaborin-4-ium-5-uide (**c**)



$C_{30}H_{39}BF_2N_6O_5$   
MW: 612.49 g/mol

To a stirring solution of BODIPY phenol (200 mg, 0.54 mmol, 1.0 eq) in THF (10 mL),  $NEt_3$  (302  $\mu$ L, 220 mg, 2.17 mmol, 4.0 eq) was added, followed by the addition of *N,N'*-di-succinimidyl carbonate (278 mg, 1.09 mmol, 2.0 eq). After stirring at rt for 16 h, TLC monitoring indicated quantitative conversion. 11-Azido-3,6,9-trioxaundecan-1-amine (480  $\mu$ L, 514 mg, 2.44 mmol, 4.5 eq) was added and stirring at rt was continued for 1 h.  $H_2O$  (150 mL) was added and the mixture was extracted with EtOAc (4 x 100 mL). The combined organic layers were dried over  $Na_2SO_4$ , filtered and the solvent was removed *in vacuo*. Flash column chromatography (*i*Hex/EtOAc = 3:1  $\rightarrow$  0:1) afforded the desired BODIPY azide **b** (156 mg, 0.26 mmol, 47%) as a dark red oil.

**$^1H$  NMR** (400 MHz,  $CD_3OD$ ):  $\delta$  (ppm) = 7.00 (s, 2H, 13-H and 15-H), 6.07 (s, 2H, 2-H and 8-H), 3.71–3.64 (m, 10H, linker- $CH_2$ ), 3.62 (t,  $^3J$  = 5.4 Hz, 2H, linker- $CH_2$ ), 3.42–3.35 (m, 4H, linker- $CH_2$ ), 2.50 (s, 6H, 1-H and 9-H), 2.13 (s, 6H, 12-H and 16-H), 1.44 (s, 6H, 3-H and 7-H).

**$^{13}C$  NMR** (102 MHz,  $^1H$ -decoupled,  $CD_3OD$ ):  $\delta$  (ppm) = 156.81 ( $C_{ar}$ ), 156.79 ( $C_{ar}$ ), 153.2 (17-C), 143.8 ( $C_{ar}$ ), 141.9 ( $C_{ar}$ ), 137.9 ( $C_{ar}$ ), 132.1 ( $C_{ar}$ ), 131.7 ( $C_{ar}$ ), 122.6 (13-C and 15-C), 122.1 (2-C and 8-C), 71.69 (linker-C), 71.67 (linker-C), 71.5 (linker-C), 71.3 (linker-C), 71.1 (linker-C), 70.8 (linker-C), 51.8 (linker-C), 42.0 (linker-C), 19.6 (12-C and 16-C), 14.6 (1-C and 9-C), 13.7 (3-C and 7-C).

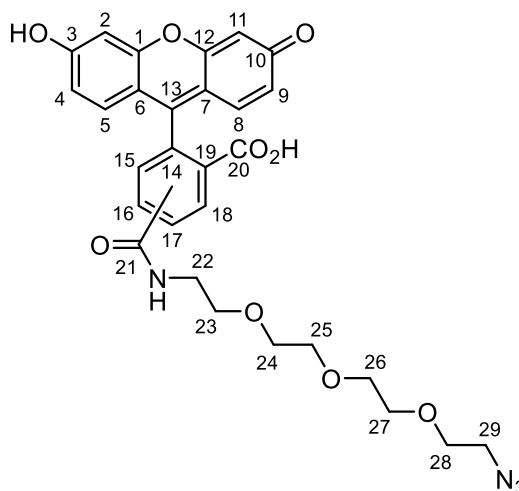
**HRESIMS**: calculated for  $C_{30}H_{39}BF_2N_6NaO_5^+$  [ $M+Na$ ] $^+$ : 635.2935, observed: 635.2934.

**UV-Vis** ( $H_2O$ ):  $\lambda_{Abs}$  (nm) = 498.

**Fluorescence** ( $H_2O$ ):  $\lambda_{Em}$  (nm) = 508.



4(5)-(2-(2-(2-(2-Azidoethoxy)ethoxy)ethoxy)carbamoyl)-2-(6-hydroxy-oxo-3*H*-xanthen-9-yl)benzoic acid (5(6)) (a)



$C_{29}H_{28}N_4O_9$   
MW: 576.55 g/mol

To a stirring solution of 5(6)-carboxyfluorescein (188.2 mg, 0.50 mmol, 1.0 eq) in DMF (1.5 mL), *N,N*-diisopropylethylamine (0.26 mL, 1.50 mmol, 3.0 eq), 4-dimethylaminopyridine (6.1 mg, 0.05 mmol, 0.1 eq), 1-(bis(dimethylamino)methylene)-1*H*-1,2,3-triazolo-[4,5-*b*]pyridinium-3-oxid hexafluorophosphate (228.1 mg, 0.60 mmol, 1.2 eq) and 11-azido-3,6,9-trioxaundecan-1-amine (120.0 mg, 0.55 mmol, 0.11 mL, 1.1 eq) were added and stirred at rt for 24 h. The solvent was removed *in vacuo* and the crude product was purified by flash column chromatography (DCM/MeOH = 9:1 → 5:1) to yield 5(6)-a (278.4 mg, 0.48 mmol, 97%) as an orange solid.

5 mg of the regioisomers 5(6)-a were separated by preparative RP-HPLC (0–50% B in 45 min) and one regioisomer **a** was used for click reactions.

**<sup>1</sup>H NMR** (400 MHz, CD<sub>3</sub>OD, 5(6)-a):  $\delta$  (ppm) = 8.44 (dd,  $^4J$  = 1.6 Hz,  $J$  = 0.7 Hz, 1H, 18-H), 8.25 (dd,  $^3J$  = 8.0 Hz,  $^4J$  = 1.6 Hz, 1H, 16-H), 8.19 (dd,  $^3J$  = 8.0 Hz,  $^4J$  = 1.4 Hz, 1H, 17-H), 8.12 (dd,  $^3J$  = 8.0 Hz,  $J$  = 0.7 Hz, 1H, 18-H), 7.69 (dd,  $^4J$  = 1.4 Hz,  $J$  = 0.8 Hz, 1H, 15-H), 7.34 (dd,  $^3J$  = 8.0 Hz,  $J$  = 0.7 Hz, 1H, 15-H), 6.75–6.71 (m, 2H, 5-H and 8-H), 6.64–6.58 (m, 2H, 2-H and 11-H), 6.54 (ddd,  $^3J$  = 8.7 Hz,  $^3J$  = 5.2 Hz,  $^4J$  = 2.4 Hz, 2H, 4-H and 9-H), 3.77–3.48 (m, 16H, linker-CH<sub>2</sub>), 2.15 (s, 1H, OH).

**<sup>13</sup>C NMR** (102 MHz, <sup>1</sup>H-decoupled, CD<sub>3</sub>OD):  $\delta$  (ppm) = 170.7 (21-C), 168.6 (20-C), 168.4 (20-C), 161.8 (3-C, 10-C), 154.3 (1-C, 12-C), 142.4 (14-C), 138.0 (13-C), 135.6 (19-C), 130.5 (5-C, 8-C), 130.3 (5-C, 8-C), 126.4 (16-C)\*, 125.9 (17-C)\*, 125.1 (18-C)\*, 124.2 (15-C)\*, 113.9 (4-C, 9-C), 111.1 (6-C, 7-C), 103.8 (2-C, 11-C), 71.83 (25-C)\*\*\*, 71.79 (26-C)\*\*\*, 71.75 (25-C)\*\*\*, 71.7 (23-C, 26-C)\*\*\*\*, 71.6 (23-C)\*\*\*, 71.5 (24-C)\*\*\*\*, 71.3 (24-C)\*\*\*\*, 71.2 (27-C)\*\*\*\*, 71.2 (27-C)\*\*\*\*, 70.6 (28-C), 70.4 (28-C), 51.9 (29-C), 51.9 (29-C), 41.4 (22-C), 41.3 (22-C).

**IR** (ATR):  $\tilde{\nu}$  (cm<sup>-1</sup>) = 3302, 3051, 2117, 1678, 1621, 1516, 1458, 1373, 1343, 1319, 1260, 1226, 1158, 1121, 1070, 982, 953, 926, 860, 837, 816, 756, 745, 720.

**HRESIMS:** calculated for  $C_{29}H_{27}N_4O_9^-$  [M-H] $^-$ : 575.1784, observed: 575.1784.

**UV-Vis** ( $H_2O$ ):  $\lambda_{Abs}$  (nm) = 495.

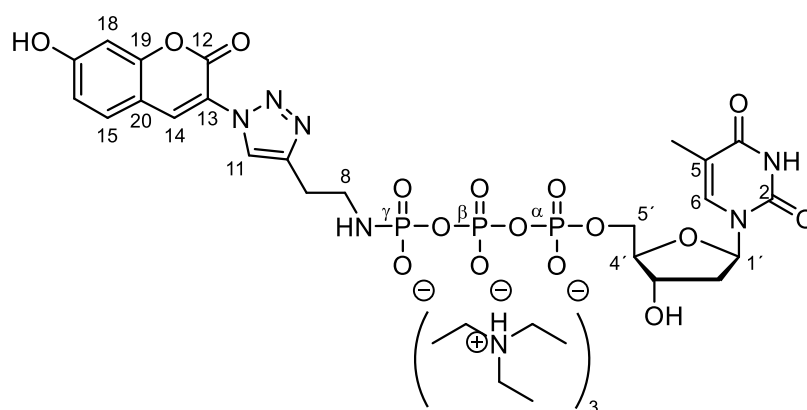
**Fluorescence** ( $H_2O$ ):  $\lambda_{Em}$  (nm) = 520.

### 5.2.11 Synthesis of $\gamma$ -Fluorophore Labeled Nucleoside Triphosphates

**General Procedure 4:** In a 0.2 mL PCR tube 8.0  $\mu$ L of the  $\gamma$ -alkyne labeled nucleoside triphosphate solution (50 mM in ddH<sub>2</sub>O, 400 nmol, 1.0 eq) was mixed with 10.4  $\mu$ L of the fluorophore azide solution (58 mM, 600 nmol, 1.5 eq) (THF solution for the coumarin **b**, BODIPY **c** and carboxyfluorescein **a**, THF/H<sub>2</sub>O = 1:1 mixture for TAMRA **e**). 0.64-1.92  $\mu$ L CuSO<sub>4</sub> solution (10 mg/mL in ddH<sub>2</sub>O, 40 nmol, 0.1-0.3 eq) and 1.60-4.80  $\mu$ L of a freshly prepared sodium ascorbate solution (100 mg/mL in H<sub>2</sub>O, 800 nmol, 2.0-6.0 eq) were added, mixed and incubated at 0 °C to 25 °C for 1-4 h (exact conditions: Table 3-6, p. 84). Purification by semipreparative RP-HPLC (0–40% B in 25 min, 40–60% B from 25-45 min) yielded  $\gamma$ -fluorophore labeled nucleotides.

**General Procedure 5:** The reactions were performed in glassware under an argon atmosphere and with magnetic stirring. The  $\gamma$ -alkyne labeled nucleoside triphosphate (1.0 eq) was dissolved in ddH<sub>2</sub>O, the fluorophore azide **b** (1.5 eq) in THF was added at rt and the brownish solution was degassed (5x). Sodium ascorbate (2.0 eq) and CuSO<sub>4</sub> (10 mol%) were added and after complete consumption of the starting material (reactions were monitored by RP-HPLC, 0–40% B in 25 min, 40–60% B from 25-45 min) the solvent was removed and the pale brown crude product was purified by preparative RP-HPLC (see below) to yield the  $\gamma$ -fluorophore labeled nucleotides.

$\gamma$ -N-(2-(1-(7-Hydroxy-2-oxo-2H-chromen-3-yl)-1H-1,2,3-triazol-4-yl)ethyl-amido)-2'-deoxythymidine-5'-triphosphate **121b**



$C_{23}H_{27}N_6O_{16}P_3$  (hypothetic free acid)  
MW: 736.42 g/mol

The synthesis was performed with **121** (16.9 mg, 20.0  $\mu$ mol, 1.0 eq), **b** (6.2 mg, 30.3  $\mu$ mol, 1.5 eq), sodium ascorbate (7.9 mg, 40.0  $\mu$ mol, 2.0 eq) and CuSO<sub>4</sub> (0.3 mg, 10 mol%) in ddH<sub>2</sub>O/THF = 1:1 (1.0 mL) at 0 °C

within 1 h as described in the general procedure **5** (chapter 5.2.11, p. 140). Purification by RP-HPLC (0–20% B in 45 min) afforded the tris-triethylammonium salt of  $\gamma$ -fluorophore labeled dTTP **121b** (14.6 mg, 14.0  $\mu$ mol, 70%) as a bright yellow solid after lyophilization.

### One pot synthesis of **121b** from dTTP

The synthesis was performed with a 100 mM dTTP solution (200  $\mu$ L, 20  $\mu$ mol, 1.0 eq) as described in the general procedure **2** (chapter 5.2.9, p. 130) within 4.5 h, using adjusted amounts of EDC·HCl and 1-aminobut-3-yne in ddH<sub>2</sub>O/DMF = 1:1 (0.8 mL). After precipitation, the crude  $\gamma$ -alkyne labeled dTTP **121** was directly used for the click reaction with fluorophore azide **b** as described in the general procedure **4** (chapter 5.2.11, p. 140) using the reagent amounts as described above. Purification by RP-HPLC (0–20% B in 45 min) afforded the tris-triethylammonium salt of the  $\gamma$ -fluorophore labeled dTTP **121b** (12.6 mg, 12.1  $\mu$ mol, 60%) as a bright yellow solid after lyophilization.

**<sup>1</sup>H NMR** (600 MHz, D<sub>2</sub>O):  $\delta$  (ppm) = 8.38 (s, 1H, 11-H), 8.36 (s, 1H, 14-H), 7.63 (d,  $^3J$  = 8.6 Hz, 1H, 15-H), 7.58 (d,  $^4J$  = 1.2 Hz, 1H, 6-H), 6.95 (dd,  $^3J$  = 8.8 Hz,  $^4J$  = 2.4 Hz, 1H, 16-H), 6.87 (d,  $^4J$  = 1.9 Hz, 1H, 18-H), 6.15 (t,  $^3J$  = 6.9 Hz, 1H, 1'-H), 4.58–4.52 (m, 1H, 3'-H), 4.18–4.05 (m, 3H, 4'-H, 5'-H), 3.36–3.25 (m, 2H, 8-H), 3.03 (t,  $^3J$  = 7.3 Hz, 2H, 9-H), 2.29–2.24 (m, 1H, 2'-H), 2.22–2.16 (m, 1H, 2''H), 1.85 (s, 3H, 7-H).

**<sup>13</sup>C NMR** (151 MHz, <sup>1</sup>H-decoupled, D<sub>2</sub>O):  $\delta$  (ppm) = 166.0 (4-C), 161.9 (12-C), 158.5 (17-C), 154.6 (19-C), 151.1 (2-C), 145.9 (13-C), 138.1 (14-C), 136.9 (6-C), 131.0 (15-C), 124.3 (11-C), 118.9 (10-C), 114.8 (16-C), 111.3 (5-C), 110.8 (20-C), 102.6 (18-C), 85.4 (d,  $^3J_{P-C}$  = 9.0 Hz, 4'-C), 84.8 (1'-C), 70.7 (3'-C), 65.2 (d,  $^2J_{P-C}$  = 6.0 Hz, 5'-C), 41.2 (8-C), 38.6 (2'-C), 27.2 (d,  $^3J_{P-C}$  = 8.9 Hz, 9-C), 11.5 (7-C).

**<sup>31</sup>P NMR** (162 MHz, <sup>1</sup>H-decoupled, D<sub>2</sub>O):  $\delta$  (ppm) = –1.78 (d,  $^2J$  = 19.0 Hz,  $\gamma$ -P), –11.81 (d,  $^2J$  = 19.2 Hz,  $\alpha$ -P), –23.04 (t,  $^2J$  = 20.5 Hz,  $\beta$ -P).

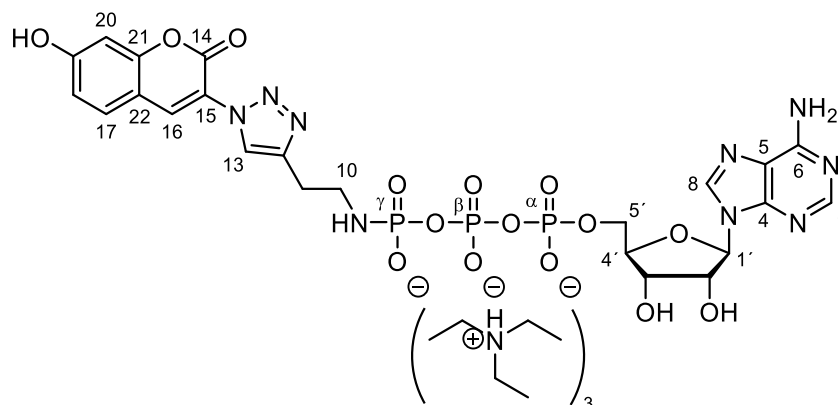
**IR** (ATR):  $\tilde{\nu}$  (cm<sup>–1</sup>) = 3331, 2987, 2692, 2500, 1728, 1695, 1651, 1605, 1475, 1419, 1398, 1327, 1222, 1116, 1082, 1052, 989, 902, 812, 798, 760, 720.

**HRESIMS**: calculated for C<sub>23</sub>H<sub>26</sub>N<sub>6</sub>O<sub>16</sub>P<sub>3</sub><sup>–</sup> [M–H]<sup>–</sup>: 735.0624, observed: 735.0620.

**UV-Vis** (H<sub>2</sub>O):  $\lambda_{Abs}$  (nm) = 393, 262.

**Fluorescence** (H<sub>2</sub>O):  $\lambda_{Em}$  (nm) = 476.

$\gamma$ -*N*-(2-(1-(7-Hydroxy-2-oxo-2*H*-chromen-3-yl)-1*H*-1,2,3-triazol-4-yl)ethyl-amido)adenosine-5'-triphosphate **122b**



$\text{C}_{28}\text{H}_{39}\text{N}_{12}\text{O}_{15}\text{P}_3$  (hypothetic free acid)  
MW: 876.61 g/mol

The synthesis was performed with **122** (8.6 mg, 10.0  $\mu\text{mol}$ , 1.0 eq), **b** (3.0 mg, 15.0  $\mu\text{mol}$ , 1.5 eq), sodium ascorbate (4.0 mg, 20.0  $\mu\text{mol}$ , 2.0 eq) and  $\text{CuSO}_4$  (0.2 mg, 10 mol%) in  $\text{ddH}_2\text{O}/\text{THF} = 1:1$  (0.5 mL) at 0 °C within 1 h as described in the general procedure **5** (chapter 5.2.11, p. 140). Purification by RP-HPLC (0–20% B in 45 min) afforded the tris-triethylammonium salt of  $\gamma$ -fluorophore labeled ATP **122b** (8.2 mg, 7.7  $\mu\text{mol}$ , 77%) as a bright yellow solid after lyophilization.

**$^1\text{H}$  NMR** (600 MHz,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = 8.28 (s, 1H, 2-H), 8.24 (s, 1H, 13-H), 8.07 (s, 1H, 16-H), 7.90 (s, 1H, 8-H), 7.40 (d,  $^3J = 8.6$  Hz, 1H, 17-H), 6.80 (dd,  $^3J = 8.6$  Hz,  $^4J = 2.2$  Hz, 1H, 18-H), 6.65 (d,  $^4J = 2.2$  Hz, 1H, 20-H), 5.91 (d,  $^3J = 5.4$  Hz, 1H, 1'-H), 4.52 (t,  $^3J = 5.2$  Hz, 1H, 2'-H), 4.46 (t,  $^3J = 4.8$  Hz, 1H, 3'-H), 4.32–4.28 (m, 1H, 4'-H), 4.27–4.15 (m, 2H, 5'-H), 3.33–3.25 (m, 2H, 10-H), 2.99 (t,  $^3J = 7.4$ , 2H, 11-H).

**$^{13}\text{C}$  NMR** (151 MHz,  $^1\text{H}$ -decoupled,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = 161.6 (14-C), 157.7 (19-C), 157.6 (6-C), 154.8 (21-C), 152.1 (2-C), 148.2 (4-C), 145.6 (15-C), 139.1 (8-C), 136.7 (16-C), 130.6 (17-C), 123.8 (13-C), 118.1 (5-C)\*, 118.0 (12-C)\*, 114.6 (18-C), 110.3 (22-C), 102.2 (20-C), 86.9 (1'-C), 83.6 (d,  $^3J_{\text{P-C}} = 8.5$  Hz, 4'-C), 74.6 (2'-C), 70.0 (3'-C), 65.0 (d,  $^2J_{\text{P-C}} = 5.6$  Hz, 5'-C), 41.2 (10-C), 27.1 (d,  $^3J_{\text{P-C}} = 8.3$  Hz, 11-C).

**$^{31}\text{P}$  NMR** (162 MHz,  $^1\text{H}$ -decoupled,  $\text{D}_2\text{O}$ ):  $\delta$  (ppm) = -1.77 (d,  $^2J = 19.9$  Hz,  $\gamma$ -P), -11.51 (d,  $^2J = 18.3$  Hz,  $\alpha$ -P), -22.95 (t,  $^2J = 17.6$  Hz,  $\beta$ -P).

**IR** (ATR):  $\tilde{\nu}$  ( $\text{cm}^{-1}$ ) = 3327, 3172, 2982, 2878, 2733, 2503, 1728, 1700, 1645, 1605, 1513, 1475, 1418, 1397, 1327, 1218, 1113, 1083, 1060, 988, 899, 846, 811, 798, 758, 718.

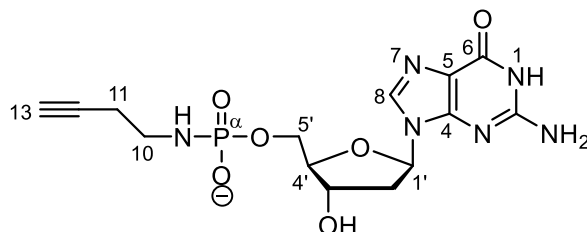
**HRESIMS**: calculated for  $\text{C}_{23}\text{H}_{25}\text{N}_9\text{O}_{15}\text{P}_3^-$  [ $\text{M-H}$ ] $^-$ : 760.0688, observed: 760.0687.

**UV-Vis** ( $\text{H}_2\text{O}$ ):  $\lambda_{\text{Abs}}$  (nm) = 354, 258.

**Fluorescence** ( $\text{H}_2\text{O}$ ):  $\lambda_{\text{Em}}$  (nm) = 478.

### 5.2.12 Synthesis of an $\alpha$ -Alkyne Labeled Nucleoside Triphosphate

$\alpha$ -*N*-(But-3-yn-1-ylamido)deoxyguanosine-5'-monophosphate **126**



$C_{14}H_{19}N_6O_6P$  (hypothetic free acid)  
MW: 398.33 g/mol

To 1.1 mL stirring solution of bis-sodium dGMP (100 mM in ddH<sub>2</sub>O, 43.0 mg, 110  $\mu$ mol, 1.0 eq) at rt were added 1.1 mL of an EDC·HCl solution (500 mM in ddH<sub>2</sub>O, 105.4 mg, 550  $\mu$ mol, 5.0 eq). After 10 min, 2.2 mL solution of 1-aminobut-3-yne (55 mM in DMF, 8.4 mg, 121  $\mu$ mol, 1.1 eq) were added and stirring continued for 6 h. The reaction was lyophilized, the dry residue was dissolved in buffer A and purified *via* RP-HPLC (0–20% B in 45 min). This afforded the triethylammonium salt of  $\alpha$ -alkyne dGMP **126** (23.4 mg, 46.8  $\mu$ mol, 43%) as a colorless solid after lyophilisation.

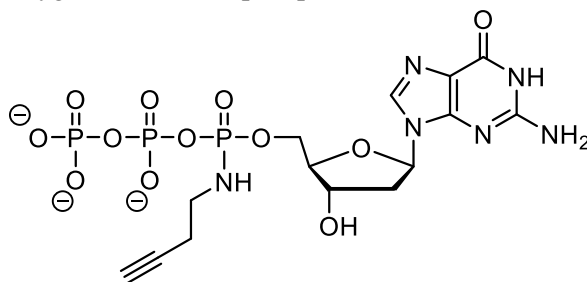
**<sup>1</sup>H NMR** (400 MHz, D<sub>2</sub>O):  $\delta$ (ppm) = 8.11 (s, 1H, 8-H), 6.34 (t,  $^3J$  = 6.5 Hz, 1H, 1'-H), 4.79–4.74 (m, 1H, 3'-H), 4.28–4.20 (m, 1H, 4'-H), 4.04–3.94 (m, 2H, 5'-H), 2.96–2.87 (m, 1H, 2'-H), 2.84–2.76 (m, 2H, 10-H), 2.63–2.55 (m, 1H, 2'-H), 2.30 (td,  $^4J$  = 2.7 Hz,  $^5J$  = 0.8 Hz, 1H, 13-H), 2.20 (td,  $^3J$  = 7.0 Hz,  $^4J$  = 2.7 Hz, 2H, 11-H).

**<sup>13</sup>C NMR** (101 MHz, <sup>1</sup>H-decoupled, D<sub>2</sub>O):  $\delta$ (ppm) = 158.8 (6-C), 153.8 (2-C), 151.4 (4-C), 137.4 (8-C), 116.3 (5-C), 85.8 (d,  $^3J_{P-C}$  = 9.2 Hz, 4'-C), 83.4 (1'-C), 83.0 (12-C), 71.2 (3'-C), 69.9 (13-C), 64.2 (d,  $^2J_{P-C}$  = 5.0 Hz, 5'-C), 40.0 (10-C), 38.2 (2'-C), 20.6 (d,  $^3J_{P-C}$  = 6.5 Hz, 11-C).

**<sup>31</sup>P NMR** (162 MHz, <sup>1</sup>H-decoupled, D<sub>2</sub>O):  $\delta$ (ppm) = 8.40 (s,  $\alpha$ -P).

**HRESIMS**: calc. for  $C_{14}H_{18}N_6O_6P^-$  [M-H]<sup>-</sup>: 397.1031; found: 397.1026.

$\alpha$ -*N*-(But-3-yn-1-ylamido)deoxyguanosine-5'-triphosphate **130**



$C_{14}H_{21}N_6O_{12}P_3$  (hypothetic free acid)  
MW: 558.27 g/mol

To a solution of the  $\alpha$ -alkyne dGMP **126** (8.3 mg, 16.6  $\mu$ mol, 1.0 eq) in dry DMF (408  $\mu$ L) at 0 °C was added DIPEA (13.6  $\mu$ L, 10.1 mg, 78.1  $\mu$ mol) and 3-methyl-1-(phenylsulfonyl)-1*H*-imidazolium triflate (**127**) (7.4 mg, 19.9  $\mu$ mol, 1.2 eq). The resulting yellow solution was incubated at 0 °C for 20 min and then added to a solution of bis-tributylammonium pyrophosphate (25.3 mg, 66.5  $\mu$ mol, 4.0 eq) in dry DMF (408  $\mu$ L) while maintaining the cooling for 1 min. After warming to rt the reaction was incubated at 25 °C, 800 rpm for 8 h using a thermomixer (EPPENDORF). The reaction was stopped by addition of diluted buffer A (50 mM, 853  $\mu$ L) and was washed with  $\text{CHCl}_3$  (3 x 2 mL). The combined aqueous layers were lyophilized and the remaining residue was dissolved in buffer A (1.0 mL), filtered and purified by RP-HPLC (0–20% B in 45 min). This afforded the tris-triethylammonium salt of the  $\alpha$ -alkyne dGTP **130** (0.26  $\mu$ mol, determined by UV-Vis, 2%) as a colorless solid.

**HRESIMS:** calc. for  $\text{C}_{14}\text{H}_{20}\text{N}_6\text{O}_{12}\text{P}_3^-$   $[\text{M}-\text{H}]^-$ : 557.0358; found: 557.0348.

### 5.3 Oligonucleotide Synthesis

All strands without artificial bases were obtained from METABION and were used without further purification. This section describes the synthesis of strands containing special bases used for experiments described within this thesis. The sequences of the strands can be found in 5.4.3, p. 148.

DNA oligonucleotide synthesis was performed on an APPLIED BIOSYSTEMS Incorporated 394 automated synthesizer. All oligodeoxynucleotides were synthesized on a 1  $\mu$ mol or 200 nmol scale, using standard DNA synthesis conditions (trityl off mode). Phosphoramidites for dA, dC, dG, dT were obtained from LINK TECHNOLOGIES. Polystyrene-based supports were purchased from GLEN RESEARCH corporation. Dry acetonitrile (<10 ppm H<sub>2</sub>O, ROTH) was used as solvent for the phosphoramidites. Standard protocols for DNA solid-phase synthesis were applied for amine and aldehyde phosphoramidites with only minor adjustments. For the phosphoramidites of amine derivatives, the capping step was omitted after addition of the special base due to protecting group problems. In addition, coupling times were doubled for all special bases.

RNA oligonucleotide synthesis was performed on an ABI 394 automated synthesizer applying A, C, G and U 2'-OTBDMS RNA phosphoramidites from LINK TECHNOLOGIES. For phosphoramidites dry acetonitrile (<10 ppm H<sub>2</sub>O, ROTH) was used as a solvent. Standard protocols for RNA solid-phase synthesis were applied for the salicylaldehyde ribophosphoramidite with only minor adjustments. Solutions (0.15 M) of the standard phosphoramidites and the salicylaldehyde ribophosphoramidite **88** and 10 equivalents of amidite per coupling were utilized. dA-CPG was used as solid support for establishing the synthesis protocol. Coupling times of 20 min were applied for the standard and 30 min for the salicylaldehyde ribophosphoramidite **88**. Monitoring of the trityl cation absorbance (498 nm) during deblocking allowed approximation of the coupling efficiency. The synthesizer was equipped with the following solutions and reagents:

Phosphoramidites: 0.10 M (0.15 M, for RNA) in acetonitrile

Activator: 0.25 M 5-[3,5-bis(trifluoromethyl)phenyl]-1*H*-tetrazole in acetonitrile (solution from SIGMA-ALDRICH)

Deblock: 3% (15% for RNA) (v/v) dichloroacetic acid in toluene

Oxidation: 25 mM iodine in 65:30:5 (v/v) acetonitrile:H<sub>2</sub>O:lutidine

Capping A: 20:50:30 (v/v) acetic anhydride:acetonitrile:lutidine

Capping B: 20% (v/v) *N*-methylimidazol in acetonitrile

### Deprotection and Purification of Oligonucleotides

DNA strands containing the aromatic amine, the aliphatic amine and the pyridine amine were cleaved off the solid support and all protecting groups were removed by incubation in 1 mL of a freshly prepared 1:1 mixture of 40% aqueous ammonia and 28% aqueous methylamine (AMA) at 65 °C for 1 h.

Hydrolysis of acetal protecting groups of DNA strands containing aldehyde bases was performed on solid support by slow washing with 2% dichloroacetic acid + 1% H<sub>2</sub>O in DCM at room temperature for 2 h. Then, the support was washed with DCM (3 x 3 mL), dried, transferred to a 2 mL reaction tube and 1 mL AMA was applied at 65 °C for 10 min.

After cooling to rt, the supernatant from the AMA deprotection was transferred to a fresh 2 mL reaction tube, the pellet was washed with 0.5 mL ddH<sub>2</sub>O and its supernatant was also transferred. The solvents were removed in a SpeedVac concentrator and the resulting pellet was dissolved in ddH<sub>2</sub>O. Analysis and purification were conducted on HPLC systems described in 5.1, p. 98. Prior to HPLC purification of DNA strands containing aldehyde bases, the crude mixture was dissolved in 10% aqueous acetic acid. Purified fractions were concentrated in a SpeedVac and lyophilized.

For deprotection of RNA strands three protocols were examined in parallel in 2 mL polypropylene reaction tubes:

- a) 1 mL saturated aqueous ammonia:EtOH = 3:1, 18 h, rt; then +10% AcOH.
- b) 1 mL 0.4 M NaOH in MeOH:H<sub>2</sub>O = 4:1, 18 h, rt; then 600 µL 1 M buffer A (5.1, p. 98).
- c) 1 mL AMA, 10 min, 65 °C.

The solvents from a)-c) were removed in a SpeedVac concentrator and the pellet was dissolved in 100 µL dry DMSO. Then 125 µL TEA·3HF were added and the mixture was incubated at 65 °C for 2.5 h. After cooling to rt, 25 µL 3 M NaOAc was added, vortexed and followed by addition of 1 mL butanol and vigorous mixing. Precipitation of the product was achieved by incubation at –70 °C for 30 min and was then sedimented by centrifugation (12500 rpm, 5 min). After removal of the supernatant, the pellet was washed twice with EtOH (absolute, 2 x 0.75 mL), dried in a SpeedVac concentrator and redissolved in 1 mL RNase free H<sub>2</sub>O. Analysis and purification were conducted on the HPLC systems described in 5.1, p. 98.

The identity of the obtained oligomers was verified by MALDI-TOF mass spectrometry and the purity was assessed by analytical RP-HPLC (5.1, p. 98).



## 5.4 Biochemical Experiments

### 5.4.1 Buffers

**Table 5-1.** List of buffers and solutions used in biochemical experiments. Note that all compositions correspond to the final working concentration. Stock solutions were prepared if necessary.

Name	Composition
Binding buffer	500 mM NaCl, 20 mM Tris-HCl, pH 8.0 at 25 °C
Binding buffer TX	500 mM NaCl, 0.05% Tween-20, 20 mM Tris-HCl, pH 8.0 at 25 °C, varying imidazole concentrations (20, 30, 200 mM)
DNA loading dye1	5% glycerine, bromophenol blue
DNA loading dye2	3.5 M urea, 15% glycerine, 44 mM Tris-HCl, 44 mM boric acid, 1 mM EDTA, bromophenol blue
Hybridization buffer1	10 mM NaCl, 10 mM Tris-HCl, pH 7.6 at 25 °C
Hybridization buffer2	100 mM NaCl, 25 mM CHES, pH 8.0 at 25 °C
Hybridization buffer3	50 mM NaCl, 10 mM CHES, pH 9.0 at 25 °C
KOD XL, pH 8	120 mM HEPES, 10 mM KCl, 6 mM (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 0.1% Triton X-100, 0.001% BSA, pH 8.0 at 25 °C
LB medium	5 g L <sup>-1</sup> yeast extract, 10 g L <sup>-1</sup> trypton, 10 g L <sup>-1</sup> NaCl,
NEB 2, pH 8	50 mM NaCl, 10 mM HEPES, 10 mM MgCl <sub>2</sub> , 1 mM DTT, pH 7.9 at 25 °C
NEB 4, pH 8	50 mM KOAc, 20 mM TrisOAc, 10 mM Mg(OAc) <sub>2</sub> , 1 mM DTT, pH 7.9 at 25 °C
Phusion buffer	50 mM KCl, 1.5 mM MgCl <sub>2</sub> 10 mM Tris-HCl, pH 8.3 at 25 °C
RNA loading dye	47.5% formamide, 0.5 mM EDTA, 0.01% SDS, 0.01% bromophenol blue
RNAP storage buffer	150 mM NaCl, 1 mM EDTA, 1 mM DTT, 20 mM NaH <sub>2</sub> PO <sub>4</sub> /Na <sub>2</sub> HPO <sub>4</sub> , pH 7.7 at 25 °C
SDS buffer	25 mM Tris, 192 mM glycine, 0.1% SDS
SDS loading dye	25 mM Tris-HCl, 5% glycerine, 1% SDS, 50 mM DTT, 0.01% bromophenol blue
SOC medium	20 g L <sup>-1</sup> vegetable peptone, 5 g L <sup>-1</sup> yeast extract, 20 mM glucose, 10 mM NaCl 10 mM MgCl <sub>2</sub> , 10 mM MgSO <sub>4</sub> , 2.5 mM KCl
T4 ligation buffer	10 mM MgCl <sub>2</sub> , 1 mM ATP, 10 mM DTT, 50 mM Tris-HCl, pH 7.5 at 25 °C
TAE buffer	40 mM Tris-OAc, 2 mM EDTA, pH 8.3 at 25 °C
TBE buffer	89 mM Tris-HCl, 89 mM boric acid, 20 mM EDTA, pH 8.0 at 25 °C
ThermoPol, pH 8.8	20 mM HEPES, 10 mM (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 10 mM KCl, 2 mM MgSO <sub>4</sub> , 0.1% Triton X-100, pH 8.8 at 25 °C
Transcription buffer	40 mM HEPES, 6 mM MgCl <sub>2</sub> , 10 mM DTT, pH 7.4 at 25 °C

### 5.4.2 Chemically Competent *E. coli* Strains

**Table 5-2.** *E. coli* strains used within this work.

<i>E. coli</i> strain	Manufacturer	Genotype
BL21 (DE3)	NEW ENGLAND BIOLABS	<i>fhuA2 ompT ΔhsdS</i> (rB <sup>-</sup> mB <sup>-</sup> ) <i>dcm<sup>+</sup> gal λ</i> (DE3) <i>i21 Δnin5</i>
NEB 5-alpha (High Efficiency)	NEW ENGLAND BIOLABS	<i>fhuA2 Φ80lacZΔM15 Δ(lacZYA-argF)</i> U169 <i>recA1 endA1 hsd R17</i> (rK <sup>-</sup> , mK <sup>+</sup> ) <i>phoA glnV44 thi<sup>-1</sup> gyrA96 relA1</i>

### 5.4.3 DNA Oligonucleotide Sequences

**Table 5-3.** Sequences of DNA strands used for mutagenesis and sequencing. Mutated positions are depicted in bold.

Name	5'-3' Sequence	Length	T <sub>M</sub> [°C]*	MW [g mol <sup>-1</sup> ]
T7_F644A_1	GGGTCCAAAGAG <b>G</b> CCGGCTTCCGTCAACAAGTGCTGG	37 mer	71.3	11440.5
T7_F644A_2	CCAGCACTTGTTGACGGAAGCCG <b>G</b> CCTCTTTGGACCC	37 mer	71.3	11302.3
T7_H784A_1	pCGCTCCTAACTTTGT <b>A</b> GCCAGCCAAGACG	29 mer	65.8	8901.8
T7_H784A_2	pATACCAGACTCCTGTTTGTGTGCATCAATCTCG	33 mer	64.2	10133.6
T7_5000_for	GCTGGGCACTAAGGCACTGGC	21 mer	59.7	6472.2
T7_5700_rev	TAGCCAGTACATCACAAGACTCA	23 mer	54.3	6985.6

\*Note that the melting temperature was calculated for 50 mM NaCl applying the nearest neighbor algorithm by the online tool: <http://www.basic.northwestern.edu/biotools/oligocalc.html> (checked July, 2015)

**Table 5-4.** Sequences of DNA strands used for transcription experiments.

Name	5'-3' Sequence	Length	MW [g mol <sup>-1</sup> ]
T1a	ATAATACGACTCACTATAGGG	21 mer	6438.3
T1	TAGTCACT <b>X</b> CTCGGGATTCCCTATAGTGAGTCGTATTAT	39 mer	
T1b	<b>X</b> = T (control)	39 mer	12023.0
T1c	<b>X</b> = S (salicylaldehyde nucleoside)	39 mer	11949.8
T1d	<b>X</b> = Ald (aldehyde nucleoside)	39 mer	11933.8
T2a	ATAATACGACTCACTATAGGCCTTTCACTACTCCTACCT	39 mer	11780.7
T2b	AGGTAGGAGTAGTGAAAGGCCTATAGTGAGTCGTATTAT	39 mer	12190.0

**Table 5-5.** Sequences of 15 mer DNA strands used for melting curve analysis.

Name	5'-3' Sequence	MW [g mol <sup>-1</sup> ]
M1	CACATT <b>A</b> XTGTTGTA	
M1a	<b>X</b> = T <sub>o</sub> (aromatic amine)	4625.9
M1b	<b>X</b> = S (salicylaldehyde nucleoside)	4553.0
M1c	<b>X</b> = Ald (aldehyde nucleoside)	4537.0
M1d	<b>X</b> = Ali (aliphatic amine)	4578.1
M1e	<b>X</b> = Pyr (pyridine amine)	4627.1
M1f	<b>X</b> = A	4566.1
M1g	<b>X</b> = C	4542.0
M2	TACAACA <b>X</b> TAATGTG	
M2a	<b>X</b> = S	4571.1
M2b	<b>X</b> = Ald	4555.1
M2c	<b>X</b> = A	4584.1
M2d	<b>X</b> = C	4560.1
M2e	<b>X</b> = G	4600.1
M2f	<b>X</b> = T	4575.1

**Table 5-6.** Sequences of DNA strands used for primer extension experiments. Artificial bases are depicted in bold.

Name	5'-3' Sequence	Length	MW [g mol <sup>-1</sup> ]
P1a	Fluo-TGGTCCGCTCGCTATAGGGAGA	23 mer	7619.0
P1b	TAGCAA <b>T</b> <sub>o</sub> TCTCCCTATAGCGAGGCGGACCA	30 mer	9244.9
P1c	TAGCAA <b>S</b> TCTCCCTATAGCGAGGCGGACCA	30 mer	9171.9
P1d	TAGCAAATCTCCCTATAGCGAGGCGGACCA	30 mer	9185.0
P1e	TAGCAACTCTCCCTATAGCGAGGCGGACCA	30 mer	9161.0

**Table 5-7.** Sequences of DNA strands used for crystallization experiments.

Name	5'-3' Sequence	Length	MW [g mol <sup>-1</sup> ]
C1a	CAC <b>T</b> <sub>o</sub> CGAGTCAGGCT	15 mer	4622.2
C1b	AGCCTGACTCG	11 mer	3317.2
C1c	AGCCTGACTCG <b>S</b>	12 mer	3617.4
C2a	ATGCGACCT <b>T</b> <sub>o</sub> TCCCT	14 mer	4243.0
C2b	AGGGAS <b>G</b> GGTC	10 mer	3104.0

#### 5.4.4 Melting Curve Experiments

Melting profiles were measured on a JASCO V-650 spectrometer using quartz glass cuvettes with 10.0 mm path length. Final samples contained 150 mM NaCl, 10 mM buffer (no amine) and 2  $\mu$ M of each strand in a final volume of 250  $\mu$ L, and depending on the experiment, additionally 1 mM amine (e.g. ethylenediamine) and 3-30  $\mu$ M CuSO<sub>4</sub>. Measurements were repeated at least three times with independent samples. Before the measurement, oligonucleotides were hybridized as a 20  $\mu$ M solution in hybridization buffer<sup>1</sup> (95 °C for 4 min followed by cooling down to 4 °C over a period of 45 min). For UV-measurements the solutions were covered with silicon oil and tightly plugged. Absorbance was recorded in the forward and reverse direction at temperatures of 5 °C to 95 °C with diverse slopes, from 0.1 to 5 K/min. At least 3 denaturing and renaturing ramps were performed and averaged for evaluation of the melting temperature.  $T_M$  values were calculated as the zero-crossing of 2<sup>nd</sup> derivate of the 349 nm background-corrected change in hyperchromicity at 260 nm.

#### 5.4.5 Site-Directed Mutagenesis

For introduction of point mutations in the T7 RNA polymerase DNA sequence, a PCR with primers containing mismatches was performed. In a total volume of 50  $\mu$ L, 50-100 ng plasmid (8 kbp, i.e. about 14 fmol) were mixed with 20 pmol forward and reverse primer in the phusion reaction buffer (Table 5-1), followed by 200  $\mu$ M dNTPs and 1 unit Phusion® HF DNA polymerase. The components were assembled on ice, mixed gently by careful pipetting and were directly transferred into a preheated thermocycler (Mastercycler Personal from EPPENDORF). Using adequately designed primers (see Table 5-3) the following temperature program was applied:

98 °C	5 min	
98 °C	15 s	30 x
60 °C	10 s	
72 °C	180 s (20 s/kbp)	
72 °C	5 min	

After PCR, methylated template plasmid was digested by adding 40 Units *DpnI* restriction endonuclease (NEW ENGLAND BIOLABS) and 8% (v/v) NEB buffer 4 (Table 5-1). After incubation at 37 °C for 1 h, 5  $\mu$ L of the mixture from the T7\_F644thA mutation was used to transform 5- $\alpha$  cells according to 5.4.6 (p. 151). Growing colonies were used to inoculate LB<sub>Carb</sub> preparatory cultures and employed for plasmid isolation (5.4.7) and subsequent sequencing (5.4.8). In case of T7\_H784A mutation, the *DpnI* reaction mixture was purified by agarose gel electrophoresis (5.4.10). The band between 8-10 kbp was cut from the gel and extracted using the NucleoSpin® gel clean-up kit (MACHEREY NAGEL) according to the manufacturer's protocol. Resulting DNA was ligated by T4 DNA ligase (NEB), applying 2 units of ligase in T4 ligation buffer (Table 5-1) at 16 °C for 45 min. Then, 5  $\mu$ L ligation reaction was used to transform 5-

alpha cells according to 5.4.6 (p. 151). Growing colonies were used to inoculate LB<sub>Carb</sub> preparatory cultures and employed for plasmid isolation (5.4.7) and subsequent sequencing (5.4.8).

#### 5.4.6 Transformation

For transformation of chemically competent *E. coli* cells, 50  $\mu$ L of BL21 (DE3) or NEB 5-alpha were thawed on ice and gently mixed with 5-100 ng vector plasmid. After 30 min incubation on ice, the cells were heat shocked for 20 s (30 s for 5-alpha) in a 42 °C water bath and then regenerated on ice for 2 min (5 min 5-alpha). Then, 250  $\mu$ L rt warm SOC medium was added (450  $\mu$ L rt warm SOC medium for 5-alpha) and the cell suspension was incubated at 37 °C, 300 rpm for 1 h. Subsequently, 20-100  $\mu$ L culture was spread on LB plates containing 100  $\mu$ g/mL carbenicillin (LB<sub>Carb</sub>) as a selection marker and incubated overnight at 37 °C. Single colonies were used to inoculate 20 mL LB<sub>Carb</sub> and incubated overnight at 37 °C, 200 rpm as preparatory cultures.

#### 5.4.7 Plasmid DNA Preparation

Plasmid DNA was isolated using the GeneJET Plasmid Miniprep Kit™ (THERMO SCIENTIFIC) according to manufacturer's instructions with one exception: The final elution of DNA from the columns was carried out in 30  $\mu$ L ddH<sub>2</sub>O.

#### 5.4.8 DNA sequencing

DNA sequencing was performed on at least 600 ng of purified plasmid by the Sanger sequencing method at the GATC BIOTECH AG. Primer sequences for the analysis of the site- directed mutagenesis can be found in Table 5-3.

#### 5.4.9 Determination of Protein and DNA Concentrations

Concentration of all mentioned macromolecules was determined by UV absorbance measurements on a NanoDrop ND-1000 Spectrophotometer (THERMO SCIENTIFIC) using 1.5  $\mu$ L sample volume.

T7 RNA polymerase protein concentration was determined by measuring the UV absorbance at 280 nm. Concentrations were calculated according to the Beer-Lambert law, using a protein sequence-based molar extinction coefficient  $\epsilon$  calculated from the EXPASYProtParam tool (<http://web.expasy.org/protparam/>, checked: July 2015) assuming all cysteines to be in a reduced state. For RNA polymerases T7 (wild-type), T7\_F644A and T7\_H784A  $\epsilon$  was calculated to 140260 M<sup>-1</sup>cm<sup>-1</sup>.

Concentration of plasmids was determined by measuring UV absorbance at 260 nm using an extinction coefficient of 0.02 (ml/ $\mu$ g) cm<sup>-1</sup>. DNA concentration from synthesized single strands was determined by measuring the UV absorbance at 260 nm. Concentrations were calculated according to Beer-Lambert law, using a DNA sequence-based molar extinction coefficient  $\epsilon$  calculated from Oligo Calc (<http://www.basic.northwestern.edu/biotools/OligoCalc.html>, checked: July 2015). For strands containing

artificial bases the extinction coefficient of their corresponding control strand was employed without corrections.

#### 5.4.10 Gel Electrophoresis of Proteins, DNA and RNA

For evaluation of the protein preparation procedure, SDS-PAGE was performed according to Laemmli<sup>[225]</sup> using discontinuous Tris gels with 5% stacking, 12% resolving gel concentration, 0.75 mm thickness and SDS loading dye (Table 5-1). Gels were run in SDS buffer at constant voltage (200 V) for 90 min and stained in Coomassie staining solution (0.1% (w/v) Coomassie Blue, 10% (v/v) acetic acid and 40% (v/v) ethanol in ddH<sub>2</sub>O) by gentle shaking at rt for several hours and then transferred to destaining solution (10% (v/v) acetic acid and 40% (v/v) ethanol in ddH<sub>2</sub>O).

Gel composition:

Component	Resolving Gel	Stacking Gel
Rotiphorese® Gel 30 (37.5:1)	8.0 mL	1.7 mL
ddH <sub>2</sub> O	6.5 mL	6.8 mL
1.5 M Tris, pH 8.8	5.0 mL	-
1.0 M Tris, pH 6.8	-	1.25 mL
10% SDS	200 µL	100 µL
10% APS	200 µL	100 µL
TEMED	20 µL	10 µL

DNA plasmids and PCR amplification products of about 8 kbp were analyzed and purified by agarose gel electrophoresis. Therefore, gels containing 1% agarose in TAE buffer and 0.01% ethidium bromide were prepared. DNA loading dye1 was added to the samples and the gel was run in TAE buffer at 100 V for 2 h. Fluorescence of the DNA fragments in the gel upon excitation at 342 nm was examined.

DNA and RNA strands of 10-30 mer oligonucleotide size were analyzed by 20% denaturing polyacrylamide gel electrophoresis on 20 x 20 cm gels of 1 mm thickness. The gels were run in TBE buffer (Table 5-1) at 40 °C applying a constant current of 40 mA per gel (maximum 1000 V) for about 3-4 h. The gel from RNA samples without fluorophores was stained with SYBR® green II for 15 min (in 0.5 x TBE) and then visualized like the fluorescein-containing samples using a LAS-3000 imaging system (RAYTEST). Gel composition: 32 mL Rotiphorese® sequencing gel concentrate, 4 mL 8 M urea, 4 mL 10 x TBE buffer, 200 µL APS, 20 µL TEMED.

#### 5.4.11 Expression and Purification of T7 RNA Polymerase Mutants

For expression of T7 RNA polymerases 1 L of LB<sub>Carb</sub> in a 5 L shaking flask was inoculated with 2 mL BL21-T7 overnight preparatory culture. The flask was incubated at 37 °C, 200 rpm until an OD<sub>600</sub> of 0.6-0.7 was reached (about 4 h). Then, IPTG was added to a final 1 mM and incubation continued at 37 °C, 200

rpm for 4 h. Cells were harvested by centrifugation (6 min, 6000 x g, 4 °C) and the pellet was frozen at –20 °C for short term storage. Pellets from protein expression were thawed and resuspended in sevenfold wet pellet weight of binding buffer T with additional 750 units endonuclease (Benzonase®, SIGMA ALDRICH) and one protease inhibitor cocktail tablet (cOmplete™, EDTA-free, ROCHE). Cells were lysed in a French pressure cell press (12000 psi; SLM AMINCO, three passes) and the lysate was centrifuged at 17000 x g for 45 min at 4 °C. The target protein was purified from the supernatant *via* His-tag affinity chromatography with a 3 mL Ni-NTA batch applying gravity flow at rt. The batch was equilibrated first with 5 column volumes (CV, i.e. here 15 mL) binding buffer, then the supernatant was loaded onto the Ni-NTA beads by stirring for 45 min at 0 °C. Subsequently, the beads were washed with 4 CV of binding buffer T, followed by 4 CV binding buffer T20. T7 RNA polymerase was eluted by adding 2.5 CV of binding buffer T200 and collected in fractions of 1 mL size, which were analyzed by UV-Vis measurements (5.4.9) and by SDS-PAGE (5.4.10). Purified protein-containing fractions were combined and the concentration was adjusted to 2.5-4.0 mg/mL by dilution or concentrating. Protein concentrating was achieved using Amicon Ultra® (MILLIPORE, regenerated cellulose) ultrafiltration devices with a 30 kDa molecular weight cut-off (MWCO) and centrifugation (up to 3200 x g). Then, the protein solution was dialyzed in a dialysis tubing (SnakeSkin™, THERMO SCIENTIFIC) against at least 100 fold volume of RNAP storage buffer (Table 5-1) at 4 °C stirring overnight. The dialyzed protein was filtered through a 0.2 µm GHP filter membrane (Acrodisc®), the concentration was adjusted to 2.5-4.0 mg/mL and DTT was added to a final concentration of 2 mM. Enzymatic activity was analyzed in a transcription assay and freezer stocks were prepared by addition of glycerol to a final 40% (v/v).

#### 5.4.12 Transcription Assay Conditions

*In vitro* transcriptions were performed in a total volume of 10-20 µL using a thermocycler (Mastercycler Personal from EPPENDORF) for temperature control. Triphosphates were thawed on ice and the components were assembled at rt.

This section describes transcription experiments for the metal base pair transcription. To the transcription buffer (Table 5-1) was added template (2 µM) and the NTPs (400 µM). Varying amounts of the subsequent components were provided with the enzyme added last:

Entry	Component	Concentration
1	Metal cation	CuSO <sub>4</sub> , MnSO <sub>4</sub> , AgNO <sub>3</sub> , FeSO <sub>4</sub> , FeCl <sub>3</sub> ·H <sub>2</sub> O
2	Amine	Ethylenediamine, methylamine, 1,2-diaminobenzene
3	Nucleotide	STP, AldTP
4	Enzyme	T7RNAP, T7RNAP_F644A, T7RNAP_H784A

Transcriptions were usually incubated at 37 °C for 5 h. Alternatively, lower (20, 25 and 30 °C) and higher (40 °C) temperatures as well as different incubation times (0.5, 1.0, 2.0, 15 h) were analyzed. After

incubation, the reaction was either stopped by heating to 70 °C for 10 min and subsequent MALDI-TOF mass spectrometry (after dialysis) or treated with RNA loading dye and applied to PAGE (5.4.10, p. 152). Remaining samples were stored at –20 °C for several days without noticeable changes.

This section describes transcription experiments involving the  $\gamma$ -alkyne and  $\gamma$ -fluorophore GTP. To transcription buffer (Table 5-1) was added the template (2  $\mu$ M) and the  $\gamma$ -alkyne or  $\gamma$ -fluorophore GTP (400  $\mu$ M). Then, 400  $\mu$ M ATP, CTP, UTP and 0-400  $\mu$ M GTP was added and transcription was started by addition of the RNA polymerase. In case no natural GTP was provided from the beginning, delayed addition after different timepoints was analyzed. Also mutant T7 RNA polymerases F644A and H784A were studied. The transcriptions were usually incubated at 37 °C for 3 h. Alternatively, lower (20 °C) temperature as well as different incubation times (0.5, 1.0, 15 h) were analyzed. After incubation, the reaction was treated as described above.

#### 5.4.13 Expression and Purification of *Bst* Pol I

*Bst* Pol I was expressed and purified as previously described.<sup>[13]</sup>

#### 5.4.14 DNA-*Bst* Pol I Co-Crystallization

Prior to crystallization experiments 500  $\mu$ M of strand CXa was annealed to strand CXb (CXc) (sequences Table 5-7) in hybridization buffer<sup>3</sup> (Table 5-1) using a thermocycler (Mastercycler Personal from EPPENDORF), applying the following temperature gradient: 95 °C for 4 min followed by cooling with 2 °C/min to 4 °C. Then *Bst* Pol I and DNA were mixed to a final 5 mg/mL and 250  $\mu$ M, respectively (about 1:3 molar ratio). Crystallization trials were performed using the vapor diffusion hanging drop technique at 18 °C with 500  $\mu$ L reservoir solution per well. For each well 0.8  $\mu$ L of reservoir solution was placed with 0.8  $\mu$ L of DNA-protein solution in one drop. Well diffracting crystals grew within 10 to 20 days using this reservoir solution: 0.1 M MES pH 5.8, 2-3% 2-methyl-2,4-pentanediol, 46-49% (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (saturated aqueous solution). Crystals were frozen in cryoprotectant solution (24% (w/v) sucrose, 55% (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (sat. aq. sol.), 2-3% 2-methyl-2,4-pentanediol, 0.1 M MES pH 5.8) and stored in liquid nitrogen until data collection.

Data were collected at the beamline PX I (Swiss Light Source (SLS), Villigen, Switzerland) and processed with the program XDS.<sup>[226]</sup> Structure solution was carried out by molecular replacement with PHASER<sup>[227]</sup> using the coordinates of 2XY5.<sup>[13]</sup> In order to reduce model bias, the temperature factors were reset to 20 for main chain and 40 for side chain and DNA atoms, respectively. Prior to model building in COOT,<sup>[228]</sup> a random shift of 0.1 Å was introduced with pdbset (CCP4. Collaborative Computational Project, Number 4. 1994. The CCP4 suite: programs for protein crystallography).<sup>[229]</sup> Restrained refinement was carried out on REFMAC5.<sup>[230]</sup> All structural figures were prepared with Pymol (SCHRÖDINGER). Data processing and refinement statistics are summarized in Table 5-8, p. 156.



---

Accession Numbers: Atomic coordinates and structure factors of the protein in complex with DNA have been deposited in the PDB at the EBI Macromolecular Structure Database (<http://www.ebi.ac.uk/pdbe>) and the code is 4uqg.

**Table 5-8.** Data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses.

	Post insertion T <sub>0</sub> -S
PDB number	4uqg
DNA strands	C2a and C2b
<b>Data collection statistics</b>	
Wavelength (Å)	1.000
Resolution range (Å)	70.16 – 2.0 (2.066 – 1.995)
Space group	P 21 21 21
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	87.442 93.996 105.416
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90
Total reflections	525616 (49867)
Unique reflections	59562 (5664)
Multiplicity	8.8 (8.8)
Completeness (%)	99.63 (96.34)
$\langle I/\sigma I \rangle$	22.05 (2.59)
<b>Refinement statistics</b>	
Wilson B-factor	30.57
$R_{\text{merge}}$	0.0846 (0.9151)
$R_{\text{meas}}$	0.08987
CC1/2	0.08987
CC*	1 (0.958)
Reflections used for R-free	
R-work	0.1844 (0.2983)
R-free	0.2165 (0.3176)
Number of non-hydrogen atoms	5208
Macromolecules	4955
Ligands	43
Water	210
Protein residues	580
RMS bonds (Å)	0.018
RMS angles (°)	1.67
Ramachandran favored (%)	98
Ramachandran outliers (%)	0.35
Clashscore	0.84
Average B-factor	35.80
Macromolecules	35.80
Ligands	48.20
Solvent	35.10

### 5.4.15 Primer Extension Experiments

Using a thermocycler (Mastercycler Personal from EPPENDORF) 20  $\mu\text{M}$  of 5'-fluorescein-labeled 23-mer primer (P1a, Table 5-6) was annealed to 50% excess of unlabeled 30-mer P1b, P1c, P1d or P1e in hybridization buffer<sup>2</sup> (Table 5-1) prior to primer extension experiments. Therefore, the following temperature gradient was applied: 95 °C for 4 min followed by cooling with 2 K/min to 4 °C.

Primer extension experiments were performed with the following polymerases: Klenow fragment, Klenow fragment *exo*<sup>-</sup>, OneTaq, Vent *exo*<sup>-</sup>, *Geobacillus stearothermophilus* Pol I (*Bst* Pol I), Terminator, Deep vent *exo*<sup>-</sup> (purchased from NEB) and KOD-XL (bought from MERCK). Modified polymerase buffers had to be used to avoid the interference of external amines with the amine-aldehyde base pairs. Primer extensions with Klenow fragment and Klenow fragment *exo*<sup>-</sup> were carried out in modified NEB buffer 2, for the polymerase KOD XL, KOD buffer was applied and Thermo buffer was used for all other polymerases (buffer composition see Table 5-1).

**General protocol:** In a 10  $\mu\text{L}$  setup 10 pmol of dsDNA was premixed in the adequate buffer (in case that two polymerases were used in one setup, both buffers were mixed). Then, 200  $\mu\text{M}$  of dNTPs and 200-400  $\mu\text{M}$  of RP-HPLC purified dSTP, dT<sub>0</sub>TP or dNaTP were added, followed by the addition of 1-2 units of polymerase. Experiments with Klenow polymerases were incubated at 37 °C, *Bst* Pol I at 60 °C and all other polymerases at 70 °C. To allow enzymatic incorporation of the artificial bases the incubation times were varied (for the exact conditions see the corresponding primer extension, chapter 3.9, 3.11.2 and 3.13). The reaction was stopped by addition of one volume of DNA loading dye<sup>2</sup> (Table 5-1) and heating to 95 °C for 5 min. Samples were analyzed by denaturing PAGE (5.4.10, p. 152).

**Mass analysis of primer extension experiments:** A 20  $\mu\text{L}$  primer extension setup containing 20 pmol of dsDNA was performed in the case mass analysis was desired. The reaction was stopped by addition of 20  $\mu\text{L}$  Rotiphenol (Roti<sup>®</sup> Phenol/Chloroform/Isoamylalcohol, ROTH) and extraction. Then, the aqueous layer (top, about 20  $\mu\text{L}$ ) was extracted twice with chloroform and concentrated *in vacuo*. Before MALDI-TOF mass spectrometry, the residue was redissolved in 3  $\mu\text{L}$  ddH<sub>2</sub>O and dialyzed on a 0.025  $\mu\text{m}$  VSWP membrane (MILLIPORE) against ddH<sub>2</sub>O for at least 4 h. 0.5  $\mu\text{L}$  from the dialyzed sample and HPA-matrix were co-crystallized on a target and then measured (also see 5.1).

## 6 Abbreviations

(d)NTP	(Deoxy)nucleoside 5'-triphosphate(s) (natural: A, C, G, T, U)
(d)XTP	(Deoxy)nucleoside 5'-triphosphate(s) (natural and artificial)
(TBA) <sub>2</sub> PP <sub>i</sub>	Bis(tri- <i>n</i> -butylammonium) pyrophosphate
°C	Degree Celsius (temperature)
A	Adenine/Adenosine (depending on context)
Å	Ångstrom, 10 <sup>-10</sup> m
Ac	Acetyl
APS	Ammonium persulfate
aq	Aqueous
BODIPY	Boron-dipyrromethene, 4,4-difluoro-4-bora-3a,4a-diaza-s-indacene
bp	Base pair
br	Broad
BuLi	Butyllithium
C	Cytosine/Cytidine (depending on context)
CD	Circular dichroism
CDI	1,1'-Carbonyldiimidazole
CHES	<i>N</i> -Cyclohexyl-2-aminoethanesulfonic acid
COSY	Correlation spectroscopy (2D NMR)
CPG	Controlled pore glass
CuAAC	Cu(I)-catalyzed alkyne azide cycloadditions
CV	Column volume
DCM	Dichloromethane
DDQ	2,3-Dichloro-5,6-dicyano-1,4-benzoquinone
DFT	Density functional theory (quantum mechanical modelling method)
DIPEA	<i>N,N</i> -Diisopropylethylamine
DMF	Dimethylformamide
DMSO	Dimethoxysulfoxide
DMTr	4,4'-Dimethoxytrityl
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
e.g.	Latin "exempli gratia", meaning: for example
EDC	1-Ethyl-3-(3-dimethyl-aminopropyl) carbodiimide
EDTA	Ethylenediaminetetraacetate
EI	Electron impact
eq	Equivalent(s)

---

ESI	Electrospray ionization
<i>et al.</i>	<i>Et alii</i> (latin = and others)
EtOAc	Ethylacetate
EtOH	Ethanol
Fm	9-Fluorenylmethyl (protecting group)
g	Gram, $10^{-3}$ kg
G	Guanine/Guanosine (depending on context)
h	Hour(s)
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine (amino acid)
HMBC	Heteronuclear multiple-bond correlation spectroscopy
HPA	Hydroxypicolinic acid
HPLC	High-pressure liquid chromatography/High-performance liquid chromatography
HR	High-resolution
HSQC	Heteronuclear single quantum coherence/correlation
Hz	Unit of frequency, $s^{-1}$
i. e.	Latin “id est”, meaning: that is
<i>i</i> Hex	Isohexane, alkane fraction with a boiling point of 40 °C at 350 mbar
<i>in vacuo</i>	In a vacuum
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
IR	Infrared (measurement), light from 700 nm to 1 mm
<i>J</i>	Coupling constant (NMR)
K	Kelvin (temperature)
KF	Klenow fragment (polymerase)
L	Liter(s)
LC	Liquid chromatography
m	Meter(s)
M	Molar ( <i>c</i> )
m	Multiplet (NMR)
<i>m/z</i>	Mass to charge ratio
MALDI-TOF	Matrix assisted laser desorption/ionization time-of-flight
<i>m</i> CPBA	<i>meta</i> -Chloroperoxybenzoic acid
MeOH	Methanol
MES	2-( <i>N</i> -Morpholino)ethanesulfonic acid
min	Minute(s)
mol	Mole(s)

---

mp	Melting point
MPD	2-Methyl-2,4-pentanediol
MS	Mass spectrometry
<i>n</i>	Normal (isomer)
Naa	Naphthalene aromatic amine (nucleoside)
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
pH	The negative decimal logarithm of the hydrogen ion activity
PhD	Doctor of philosophy (university degree)
$pK_a$	Logarithmic acid dissociation constant $K_a$ , $pK_a = -\log_{10} K_a$
ppm	Parts per million
Py	Pyridine
q	Quartet (NMR)
$R_f$	Retention factor (thin-layer chromatography)
RNA	Ribonucleic acid
RP	Reversed-phase
$R_t$	Retention time (chromatography)
rt	Room temperature
S	Salicylaldehyde (nucleoside)
s	Second (unit of time)
s	Singlet (NMR)
SDS	Sodium dodecyl sulfate
SNI	Single nucleotide insertion
T	Thymine/Thymidine (depending on context)
t	Triplet (NMR)
<i>t</i>	Tertiary (isomer)
TAMRA	Tetramethylrhodamine
TBA	Tetrabutylamine/Tetrabutylammonium
TBAF	Tetrabutylammoniumfluorid
TBS/TBDMS	<i>tert</i> -Butyldimethylsilyl
TBTA	Tris[(1-benzyl-1 <i>H</i> -1,2,3-triazol-4-yl)methyl]amine
TEA	Triethylamine/Triethylammonium
TEMED	Tetramethylethylenediamine
TES	Triethylsilyl

---

TFA	Trifluoroacetic (acid)/Trifluoroacetyl
THF	Tetrahydrofuran
THPTA	Tris[(1-hydroxypropyl-1 <i>H</i> -1,2,3-triazol-4-yl)methyl]amine
TIPS	Triisopropylsilyl
TLC	Thin-layer chromatography
T <sub>M</sub>	Melting temperature (DNA double strand)
T <sub>o</sub>	Aromatic amine (nucleoside)
Tol	Toluene
TOM	[(Triisopropylsilyl)oxy]methyl
Tris	Tris(hydroxymethyl)aminomethane
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
UV	Ultraviolet range of light, 10-380 nm
v/v	Volume per volume
vis	Visible range of light, 380-700 nm
w/v	Weight per volume
$\delta$	Chemical shift (NMR)

## 7 Appendix

### 7.1 Protein Sequences

**T7 RNA polymerase** protein sequence (without linker and tag); 883 amino acids, the mutated residues are marked bold:

mntiniaknd fsdielaaip fntladhyge rlareqlale hesyemgear frkmferqlk  
agevadnaaa kplittllpk miarindwfe evkakragrp tafqlqeik peavayitik  
ttlacitsad nttvqavasa igraiedear fgrirdleak hfknveeq nkrvghvykk  
afmqvveadm lskglggee wsswhkeds hvgvrcieml iestgmvsllh rqnagvvvgd  
setielapey aeaiatruga lagispmfqp cvvppkpwgt itgggywang rrplalvrth  
skkalmryed vympevykai niaqntawki nkkvlavanv itkwkhpve dipaiereel  
pmkpedidmn pealtawkra aaavyrkdka rksrrislef mleqankfan hkaiwfpynm  
dwrgrvyavs mfnppgndmt kglitlakgk pigkegyywl kihgancagv dkvpfperik  
fieenhenim acaksplent wwaeqdsfpc flafcfeyag vqhhglsync slplafdgsc  
sgiqhfsaml rdevggravn llpsetvqdi ygivakkvne ilqadaingt dnevvvtvde  
ntgeisekyk lgtkalagqw laygvtrsvt krsvmtlayg skefgfrqqv ledtiqpaid  
sgkglmftqp nqaagymakl iwesvsvtvv aaveamnwlk saakllaaev kdkktgeilr  
krcavhwvtp dgfpvwqeyk kpiqtrlnlm flgqfrlqpt intnkdeid ahkqesgiap  
nfvhsqdgsh lrktvvwahe kygiesfali hdsfgtipad aanlfkavre tmvdtyescd  
vladydqfa dqlhesqldk mpalpakgnl nlrdelesdf afa

### 7.2 DNA Sequences

**T7 RNA polymerase** gene sequence (without linker and tag):

atgaacacgattaacatcgctaagaacgacttctctgacatcgaactggctgctatcccgttaacactctggctgaccattacgggtgagcgtttagctcgcg  
aacagttggcccttgagcatgagctcttacgagatgggtgaagcacgcttcgcaagatgtttgagcgtcaactaaagctgggtgaggtgaggataacgctg  
ccgccaagcctctcatcactaccctactccctaagatgattgcacgcatcaacgactggtttgaggaagtgaagctaaagcgcggcaagcggcgacagc  
cttcagttcctgcaagaaatcaagccggaagccgtagcgtacatcaccattaagaccactctggcttgccctaaccagtctgacaatacaaccgttcaggc  
tgtagcaagcgcgaatcggtcgccgcatgaggacgaggtcgcctcggtgctatccgtgacctgaagctaagcacttcaagaaaaacgttgaggaaaca  
ctcaacaagcgcgtagggcacgtctacaagaaagcatttatcaagttgctgaggtgacatgctcttaagggtctactcgggtggcaggcgtggtcttcg  
tggcataaggaagactctattcatgtaggagtagctgcatcgagatgctcattgagtcacccggaatggtagcttacaccgccaaaatgctggcgtagta  
ggtcaagactctgagactatcgaaactgcacctgaatagctgaggctatcgcaaccgctgcaggtgcgctggctggcatctctccagtgtccaaccttgc  
gtagttcctcctaagccgtggactggcattactgggtggctattgggctaacggctgctgctccttggcgctgggtgctactcacagtaagaaagcactga  
tgcgctacgaagacgtttacatgctgaggtgtacaaagcattaacattgcgcaaaacaccgcatggaaaatcaacaagaaagtcctagcgggtcgccaa  
cgtaatcaccaagtgaagcattgtccggtcaggacatccctgcgattgagcgtgaagaactcccgatgaaccggaagacatcgacatgaatcctgag  
gctctaccgctggaaacgtgctgccgtgctgtgtaccgcaaggacaaggctcgcaagctcgcggtatcagccttgagttcatgcttgagcaagccaa  
taagtttgtaaccataaggccatctggtcccttacaacatggactggcgcggtcgtgtttacgctgtgtcaatgttcaaccgcaaggtaacgatatgacca  
aaggactgcttacgctggcgaaaggtaaaccaatcggttaagggaaggttactactggctgaaaatccacgggtgcaactgtcggggtgctgataagggtcc  
gttccctgagcgcacatgaagtcattgaggaaaaccacgagaacatcatggttcgctaaagtctccactggagaacacttgggtgggctgagcaagattctcc  
gttctgcttccttgctgttctgctttgagtagctgggttacagcaccacggcctgagctataactgctccctccgctggcggttgacgggtcttgcctggcat  
ccagcacttctccgcatgctccgagatgaggtaggtggtcgcggttaactgcttcctagtgaaccgttcaggacatctacgggattgttgtaagaaa  
gtcaacgagattctacaagcagacgcaatcaatgggaccgataacgaagtagttaccgtgaccgatgagaacactgggtgaaatctctgagaaagtcaagc  
tgggactaaaggcactggctggtcaatggctggcttacgggtgtactcgagtgactaagcgttcagtcacgctggcttacgggtccaaagagttcg  
gcttcctgaacaagtgtggaagataccattcagccagctattgattccggcaagggtctgatgttactcagccgaatcaggctgctggatacatggctaa  
gctgatttgggaatctgtgagcgtgacggtgtagctgcggttaagcaatgaactggcttaagtctgctgtaagctgctggctgctgaggtcaagataa  
gaagactggagagattctcgcaagcgttgctgtgcatgggtaactcctgatgtttccctgtgtggcaggaataacaagaagcctattcagacgcgcttg  
aacctgatgttcctcggtcagttccgttacagcctaccattaacaccaacaaagatagcgagattgatgcacacaaacaggagctgtgtatcgtcctaact  
ttgtacacagccaagacggtagccaccttctgaagactgtagtgtggcacacgagaagtaggaatgaatctttgactgattcacgactccttcgggtac  
cattccggctgacgctgcgaacctgttcaaacgagtcgcgcaaaactatggttgacacatatgagttgtgatgtactggctgatttctacgaccagttcgctg  
accagttgcacgagttcaattggacaaaatgccagcacttccggctaaaggtaactgaacctccgtgacatcttagagtcggacttcgcgttcgctaa



## 8 References

- [1] V. Vivet-Boudou, J. Didierjean, C. Isel, R. Marquet, *Cell. Mol. Life Sci.* **2006**, 63, 163-186. *Nucleoside and Nucleotide Inhibitors of Hiv-1 Replication.*
- [2] N. M. F. S. A. Cerqueira, P. A. Fernandes, M. J. Ramos, *Chem. Eur. J.* **2007**, 13, 8507-8515. *Understanding Ribonucleotide Reductase Inactivation by Gemcitabine.*
- [3] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, *et al.*, *Science* **2009**, 323, 133-138. *Real-Time DNA Sequencing from Single Polymerase Molecules.*
- [4] D. A. Malyshev, F. E. Romesberg, *Angew. Chem. Int. Ed.* **2015**, 54, 11930-11944. *The Expanded Genetic Alphabet.*
- [5] L. Li, M. Degardin, T. Lavergne, D. A. Malyshev, K. Dhami, P. Ordoukhanian, F. E. Romesberg, *J. Am. Chem. Soc.* **2014**, 136, 826-829. *Natural-Like Replication of an Unnatural Base Pair for the Expansion of the Genetic Alphabet and Biotechnology Applications.*
- [6] V. B. Pinheiro, A. I. Taylor, C. Cozens, M. Abramov, M. Renders, S. Zhang, J. C. Chaput, J. Wengel, S.-Y. Peak-Chew, S. H. McLaughlin, *et al.*, *Science (New York, N.Y.)* **2012**, 336, 341-344. *Synthetic Genetic Polymers Capable of Heredity and Evolution.*
- [7] J. Korlach, A. Bibillo, J. Wegener, P. Peluso, T. T. Pham, I. Park, S. Clark, G. A. Otto, S. W. Turner, *Nucleosides Nucleotides Nucl. Acids* **2008**, 27, 1072-1083. *Long, Processive Enzymatic Dna Synthesis Using 100% Dye-Labeled Terminal Phosphate-Linked Nucleotides.*
- [8] K. Dhami, D. A. Malyshev, P. Ordoukhanian, T. Kubelka, M. Hocek, F. E. Romesberg, *Nucleic Acids Res.* **2014**, 42, 10235-10244. *Systematic Exploration of a Class of Hydrophobic Unnatural Base Pairs Yields Multiple New Candidates for the Expansion of the Genetic Alphabet.*
- [9] Y. J. Seo, G. T. Hwang, P. Ordoukhanian, F. E. Romesberg, *J. Am. Chem. Soc.* **2009**, 131, 3246-3252. *Optimization of an Unnatural Base Pair Towards Natural-Like Replication.*
- [10] M. Kimoto, R. Kawai, T. Mitsui, S. Yokoyama, I. Hirao, *Nucleic Acids Res.* **2009**, 37, e14-e14. *An Unnatural Base Pair System for Efficient Pcr Amplification and Functionalization of DNA Molecules.*
- [11] J. C. Morales, E. T. Kool, *J. Am. Chem. Soc.* **1999**, 121, 2323-2324. *Minor Groove Interactions between Polymerase and DNA: More Essential to Replication Than Watson-Crick Hydrogen Bonds?*
- [12] G. H. Clever, K. Polborn, T. Carell, *Angew. Chem. Int. Ed.* **2005**, 44, 7204-7208. *A Highly DNA-Duplex-Stabilizing Metal-Salen Base Pair.*
- [13] C. Kaul, M. Muller, M. Wagner, S. Schneider, T. Carell, *Nat. Chem.* **2011**, 3, 794-800. *Reversible Bond Formation Enables the Replication and Amplification of a Crosslinking Salen Complex as an Orthogonal Base Pair.*
- [14] S. M. Hacker, D. Pagliarini, T. Tischer, N. Hardt, D. Schneider, M. Mex, T. U. Mayer, M. Scheffner, A. Marx, *Angew. Chem. Int. Ed.* **2013**, 52, 11916-11919. *Fluorogenic Atp Analogues for Online Monitoring of Atp Consumption: Observing Ubiquitin Activation in Real Time.*
- [15] J. L. T. J. M. Berg, L. Stryer, *Biochemistry*, 5th ed., W. H. Freeman and Company, New York, **2002**.
- [16] R. Dahm, *Hum. Genet.* **2008**, 122, 565-581. *Discovering DNA: Friedrich Miescher and the Early Years of Nucleic Acid Research.*
- [17] P. A. Levene, *J. Biol. Chem.* **1919**, 40, 415-424. *The Structure of Yeast Nucleic Acid: Iv. Ammonia Hydrolysis.*
- [18] W. Klein, S. J. Thannhauser, in *Hoppe-Seyler's Zeitschrift für physiologische Chemie*, Vol. 231, **1935**, p. 96.
- [19] D. H. Hayes, A. M. Michelson, A. R. Todd, *Journal of the Chemical Society (Resumed)* **1955**, 808-815. *Nucleotides. Part Xxx. Mononucleotides Derived from Deoxyadenosine and Deoxyguanosine.*
- [20] A. R. Todd, *J. Chem. Soc.* **1946**, 647-653. *Synthesis in the Study of Nucleotides.*

- [21] O. T. Avery, C. M. MacLeod, M. McCarty, *J. Exp. Med.* **1944**, 79, 137-158. *Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii.*
- [22] J. D. Watson, F. H. Crick, *Nature* **1953**, 171, 737-738. *Molecular Structure of Nucleic Acids.*
- [23] M. H. F. Wilkins, A. R. Stokes, H. R. Wilson, *Nature* **1953**, 171, 738-740. *Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids.*
- [24] J. D. Watson, F. H. C. Crick, *Nature* **1953**, 171, 964-967. *Genetical Implications of the Structure of Deoxyribonucleic Acid.*
- [25] M. Meselson, F. W. Stahl, *Proc. Natl. Acad. Sci. USA* **1958**, 44, 671-682. *The Replication of DNA in Escherichia Coli.*
- [26] M. Nirenberg, P. Leder, *Science* **1964**, 145, 1399-1407. *Rna Codewords and Protein Synthesis the Effect of Trinucleotides Upon the Binding of Srna to Ribosomes.*
- [27] M. W. Nirenberg, J. H. Matthaei, *Proc. Natl. Acad. Sci. USA* **1961**, 47, 1588-1602. *The Dependence of Cell-Free Protein Synthesis in E. Coli Upon Naturally Occurring or Synthetic Polyribonucleotides.*
- [28] R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. E. Dickerson, *Nature* **1980**, 287, 755-758. *Crystal Structure Analysis of a Complete Turn of B-DNA.*
- [29] A. J. Bruce Alberts, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter, *Molecular Biology of the Cell*, 4th ed., Garland Science, New York, **2002**.
- [30] E. A. Kowal, M. Ganguly, P. S. Pallan, L. A. Marky, B. Gold, M. Egli, M. P. Stone, *J. Phys. Chem. B* **2011**, 115, 13925-13934. *Altering the Electrostatic Potential in the Major Groove: Thermodynamic and Structural Characterization of 7-Deaza-2'-Deoxyadenosine:Dt Base Pairing in DNA.*
- [31] V. Amarnath, A. D. Broom, *Chem. Rev.* **1977**, 77, 183-217. *Chemical Synthesis of Oligonucleotides.*
- [32] W. T. Markiewicz, M. Wiewiórowski, *Nucleic Acids Res.* **1978**, 1, s185-s190. *A New Type of Silyl Protecting Groups in Nucleoside Chemistry.*
- [33] J. Ludwig, F. Eckstein, *J. Org. Chem.* **1989**, 54, 631-635. *Rapid and Efficient Synthesis of Nucleoside 5'-O-(1-Thiotriphosphates), 5'-Triphosphates and 2',3'-Cyclophosphorothio-ates Using 2-Chloro-4h-1,3,2-Benzodioxaphosphorin-4-One.*
- [34] U. Niedballa, H. Vorbrüggen, *Angew. Chem. Int. Ed. Engl.* **1970**, 9, 461-462. *A General Synthesis of Pyrimidine Nucleosides.*
- [35] C. P. Ashcroft, Y. Dessi, D. A. Entwistle, L. C. Hesmondhalgh, A. Longstaff, J. D. Smith, *Org. Process Res. Dev.* **2012**, 16, 470-483. *Route Selection and Process Development of a Multikilogram Route to the Inhaled A2a Agonist Uk-432,097.*
- [36] R. Bihovsky, C. Selick, I. Giusti, *J. Org. Chem.* **1988**, 53, 4026-4031. *Synthesis of C-Glucosides by Reactions of Glucosyl Halides with Organocuprates.*
- [37] S. De Bernardo, M. Weigle, *J. Org. Chem.* **1976**, 41, 287-290. *Synthesis of the Pyrazomycins.*
- [38] U. Wichai, S. A. Woski, *Org. Lett.* **1999**, 1, 1173-1175. *Disiloxane-Protected 2-Deoxyribonolactone as an Efficient Precursor to 1,2-Dideoxy-1-B-Aryl-D-Ribofuranoses.*
- [39] M. Štefko, L. Slavětinská, B. Klepetářová, M. Hocek, *J. Org. Chem.* **2009**, 75, 442-449. *A General and Efficient Synthesis of Pyridin-2-Yl C-Ribonucleosides Bearing Diverse Alkyl, Aryl, Amino, and Carbamoyl Groups in Position 6.*
- [40] H. C. Zhang, G. D. Daves, *J. Org. Chem.* **1992**, 57, 4690-4696. *Syntheses of 2'-Deoxypseudouridine, 2'-Deoxyformycin B, and 2',3'-Dideoxyformycin B by Palladium-Mediated Glycal-Aglycon Coupling.*
- [41] M. Tomás-Gamasa, S. Serdjukow, M. Su, M. Müller, T. Carell, *Angew. Chem. Int. Ed.* **2015**, 54, 796-800. *"Post-It" Type Connected DNA Created with a Reversible Covalent Cross-Link.*
- [42] Y. Wu, A. K. Ogawa, M. Berger, D. L. McMinn, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **2000**, 122, 7621-7632. *Efforts toward Expansion of the Genetic Alphabet: Optimization of Interbase Hydrophobic Interactions.*

- [43] B. M. Trost, L. S. Kallander, *J. Org. Chem.* **1999**, 64, 5427-5435. *A Versatile Enantioselective Strategy toward L-C-Nucleosides: A Total Synthesis of L-Showdomycin.*
- [44] S. Matsuda, F. E. Romesberg, *J. Am. Chem. Soc.* **2004**, 126, 14419-14427. *Optimization of Interstrand Hydrophobic Packing Interactions within Unnatural DNA Base Pairs.*
- [45] S. Törnroth-Horsefield, R. Neutze, *Proc. Natl. Acad. Sci. USA* **2008**, 105, 19565-19566. *Opening and Closing the Metabolite Gate.*
- [46] P. J. Lane, J. A. Ledbetter, F. M. McConnell, K. Draves, J. Deans, G. L. Schieven, E. A. Clark, *J. Immunol.* **1991**, 146, 715-722. *The Role of Tyrosine Phosphorylation in Signal Transduction through Surface Ig in Human B Cells. Inhibition of Tyrosine Phosphorylation Prevents Intracellular Calcium Release.*
- [47] C. H. Wong, S. L. Haynie, G. M. Whitesides, *J. Am. Chem. Soc.* **1983**, 105, 115-117. *Preparation of a Mixture of Nucleoside Triphosphates from Yeast Rna: Use in Enzymic Synthesis Requiring Nucleoside Triphosphate Regeneration and Conversion to Nucleoside Diphosphate Sugars.*
- [48] J. Ludwig, *Acta Biochim. Biophys. Acad. Sci. Hung.* **1981**, 16, 131-133. *A New Route to Nucleoside 5'-Triphosphates.*
- [49] J. L. Ruth, Y. C. Cheng, *Mol. Pharmacol.* **1981**, 20, 415-422. *Nucleoside Analogues with Clinical Potential in Antivirus Chemotherapy. The Effect of Several Thymidine and 2'-Deoxycytidine Analogue 5'-Triphosphates on Purified Human (Alpha, Beta) and Herpes Simplex Virus (Types 1, 2) DNA Polymerases.*
- [50] D. E. Hoard, D. G. Ott, *J. Am. Chem. Soc.* **1965**, 87, 1785-1788. *Conversion of Mono- and Oligodeoxyribonucleotides to 5'-Triphosphates I.*
- [51] J. Caton-Williams, M. Smith, N. Carrasco, Z. Huang, *Org. Lett.* **2011**, 13, 4156-4159. *Protection-Free One-Pot Synthesis of 2'-Deoxynucleoside 5'-Triphosphates and DNA Polymerization.*
- [52] J. Caton-Williams, R. Hoxhaj, B. Fiaz, Z. Huang, *Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.]* **2013**, 0 1, Unit-1.30. *Use of a Novel 5'-Regioselective Phosphitylating Reagent for One-Pot Synthesis of Nucleoside 5'-Triphosphates from Unprotected Nucleosides.*
- [53] S. Mohamady, A. Desoky, S. D. Taylor, *Org. Lett.* **2012**, 14, 402-405. *Sulfonyl Imidazolium Salts as Reagents for the Rapid and Efficient Synthesis of Nucleoside Polyphosphates and Their Conjugates.*
- [54] G. S. Cremonnik, A. Hofer, H. J. Jessen, *Angew. Chem. Int. Ed.* **2014**, 53, 286-289. *Iterative Synthesis of Nucleoside Oligophosphates with Phosphoramidites.*
- [55] M. Yoshikawa, T. Kato, T. Takenishi, *Tetrahedron Lett.* **1967**, 8, 5065-5068. *A Novel Method for Phosphorylation of Nucleosides to 5'-Nucleotides.*
- [56] F. Cramer, M. Winter, *Chem. Ber.* **1961**, 94, 989-996. *Imidoester, Vi. Katalytische Wirkung Von Dimethylformamid Bei Reaktionen Von Phosphorsäureester-Chloriden.*
- [57] S. K. Jarchow-Choy, A. T. Krueger, H. Liu, J. Gao, E. T. Kool, *Nucleic Acids Res.* **2011**, 39, 1586-1594. *Fluorescent Xdna Nucleotides as Efficient Substrates for a Template-Independent Polymerase.*
- [58] T. Ohtsuki, M. Kimoto, M. Ishikawa, T. Mitsui, I. Hirao, S. Yokoyama, *Proc. Natl. Acad. Sci. USA* **2001**, 98, 4922-4925. *Unnatural Base Pairs for Specific Transcription.*
- [59] H. A. Staab, H. Schaller, F. Cramer, *Angew. Chem.* **1959**, 71, 736-736. *Imidazolid Der Phosphorsäure.*
- [60] L. Goldman, J. W. Marsico, G. W. Anderson, *J. Am. Chem. Soc.* **1960**, 82, 2969-2970. *The Preparation of Adenosine-5' Imidazol-1-Ylphosphonate and Its Reactions with Nucleophiles. A Novel Synthesis of Nucleotide Coenzymes.*
- [61] M. Maeda, A. D. Patel, A. Hampton, *Nucleic Acids Res.* **1977**, 4, 2843-2853. *Formation of Ribonucleotide 2', 3'-Cyclic Carbonates During Conversion of Ribonucleoside 5'-Phosphates to Diphosphates and Triphosphates by the Phosphorimidazolidate Procedure.*
- [62] J. E. Marugg, M. Tromp, E. Kuyl-Yeheskiely, G. A. van der Marel, J. H. van Boom, *Tetrahedron Lett.* **1986**, 27, 2661-2664. *A Convenient and General Approach to the Synthesis of Properly Protected D-Nucleoside-3'-Hydrogenphosphonates Via Phosphite Intermediates.*

- [63] Z. Yang, A. M. Sismour, S. A. Benner, *Nucleic Acids Res.* **2007**, *35*, 3118-3127. *Nucleoside Alpha-Thiotriphosphates, Polymerases and the Exonuclease Iii Analysis of Oligonucleotides Containing Phosphorothioate Linkages.*
- [64] J. Caton-Williams, L. Lin, M. Smith, Z. Huang, *Chem. Commun.* **2011**, *47*, 8142-8144. *Convenient Synthesis of Nucleoside 5[Prime or Minute]-Triphosphates for Rna Transcription.*
- [65] H. J. Jessen, N. Ahmed, A. Hofer, *Org. Biomol. Chem.* **2014**, *12*, 3526-3530. *Phosphate Esters and Anhydrides—Recent Strategies Targeting Nature's Favoured Modifications.*
- [66] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2001**, *40*, 2004-2021. *Click Chemistry: Diverse Chemical Function from a Few Good Reactions.*
- [67] L. Liang, D. Astruc, *Coord. Chem. Rev.* **2011**, *255*, 2933-2945. *The Copper(I)-Catalyzed Alkyne-Azide Cycloaddition (Cuaac) "Click" Reaction and Its Applications. An Overview.*
- [68] A. Michael, *J. prakt. Chem.* **1893**, *48*, 94-95. *Ueber Die Einwirkung Von Diazobenzolimid Auf Acetylendicarbonsäuremethylester.*
- [69] R. Huisgen, *Proc. Chem. Soc.* **1961**, 357-396. *Proceedings of the Chemical Society. October 1961.*
- [70] R. Huisgen, *Angew. Chem. Int. Ed. Engl.* **1963**, *2*, 565-598. *1,3-Dipolar Cycloadditions. Past and Future.*
- [71] V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2002**, *41*, 2596-2599. *A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes.*
- [72] C. W. Tornøe, C. Christensen, M. Meldal, *J. Org. Chem.* **2002**, *67*, 3057-3064. *Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides.*
- [73] V. O. Rodionov, S. I. Presolski, D. Díaz Díaz, V. V. Fokin, M. G. Finn, *J. Am. Chem. Soc.* **2007**, *129*, 12705-12712. *Ligand-Accelerated Cu-Catalyzed Azide-Alkyne Cycloaddition: A Mechanistic Report.*
- [74] J. E. Hein, V. V. Fokin, *Chem. Soc. Rev.* **2010**, *39*, 1302-1315. *Copper-Catalyzed Azide-Alkyne Cycloaddition (Cuaac) and Beyond: New Reactivity of Copper(I) Acetylides.*
- [75] F. Himo, T. Lovell, R. Hilgraf, V. V. Rostovtsev, L. Noodleman, K. B. Sharpless, V. V. Fokin, *J. Am. Chem. Soc.* **2005**, *127*, 210-216. *Copper(I)-Catalyzed Synthesis of Azoles. Dft Study Predicts Unprecedented Reactivity and Intermediates.*
- [76] M. Ahlquist, V. V. Fokin, *Organometallics* **2007**, *26*, 4389-4391. *Enhanced Reactivity of Dinuclear Copper(I) Acetylides in Dipolar Cycloadditions.*
- [77] S. Sun, P. Wu, *J. Phys. Chem. A* **2010**, *114*, 8331-8336. *Mechanistic Insights into Cu(I)-Catalyzed Azide-Alkyne "Click" Cycloaddition Monitored by Real Time Infrared Spectroscopy.*
- [78] V. O. Rodionov, V. V. Fokin, M. G. Finn, *Angew. Chem. Int. Ed.* **2005**, *44*, 2210-2215. *Mechanism of the Ligand-Free Cu(I)-Catalyzed Azide-Alkyne Cycloaddition Reaction.*
- [79] Q. Wang, T. R. Chan, R. Hilgraf, V. V. Fokin, K. B. Sharpless, M. G. Finn, *J. Am. Chem. Soc.* **2003**, *125*, 3192-3193. *Bioconjugation by Copper(I)-Catalyzed Azide-Alkyne [3 + 2] Cycloaddition.*
- [80] T. R. Chan, R. Hilgraf, K. B. Sharpless, V. V. Fokin, *Org. Lett.* **2004**, *6*, 2853-2855. *Polytriazoles as Copper(I)-Stabilizing Ligands in Catalysis.*
- [81] L. Ciavatta, D. Ferri, R. Palombi, *J. Inorg. Nucl. Chem.* **1980**, *42*, 593-598. *On the Equilibrium  $\text{Cu}^{2+} + \text{Cu}(S) \rightleftharpoons 2\text{Cu}^{+}$ .*
- [82] V. Hong, S. I. Presolski, C. Ma, M. G. Finn, *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 9879-9883. *Analysis and Optimization of Copper-Catalyzed Azide-Alkyne Cycloaddition for Bioconjugation.*
- [83] C. Besanceney-Webler, H. Jiang, T. Zheng, L. Feng, D. Soriano del Amo, W. Wang, L. M. Klivansky, F. L. Marlow, Y. Liu, P. Wu, *Angew. Chem. Int. Ed.* **2011**, *50*, 8051-8056. *Increasing the Efficacy of Bioorthogonal Click Reactions for Bioconjugation: A Comparative Study.*
- [84] D. S. d. Amo, W. Wang, H. Jiang, C. Besanceney, A. Yan, M. Levy, Y. Liu, F. L. Marlow, P. Wu, *J. Am. Chem. Soc.* **2010**, *132*, 16893-16899. *Biocompatible Copper(I) Catalysts for in Vivo Imaging of Glycans.*

- [85] S. Lal, S. Díez-González, *J. Org. Chem.* **2011**, 76, 2367-2373. *[Cubr(Pph3)3] for Azide–Alkyne Cycloaddition Reactions under Strict Click Conditions.*
- [86] E. Paredes, S. R. Das, *Bioorg. Med. Chem. Lett.* **2012**, 22, 5313-5316. *Optimization of Acetonitrile Co-Solvent and Copper Stoichiometry for Pseudo-Ligandless Click Chemistry with Nucleic Acids.*
- [87] A. H. El-Sagheer, T. Brown, *Chem. Soc. Rev.* **2010**, 39, 1388-1405. *Click Chemistry with DNA.*
- [88] M. Haque, X. Peng, *Sci. China Chem.* **2014**, 57, 215-231. *DNA-Associated Click Chemistry.*
- [89] P. M. E. Gramlich, C. T. Wirges, A. Manetto, T. Carell, *Angew. Chem. Int. Ed.* **2008**, 47, 8350-8358. *Postsynthetic DNA Modification through the Copper-Catalyzed Azide–Alkyne Cycloaddition Reaction.*
- [90] C. J. Burrows, J. G. Muller, *Chem. Rev.* **1998**, 98, 1109-1152. *Oxidative Nucleobase Modifications Leading to Strand Scission.*
- [91] J. Gierlich, G. A. Burley, P. M. E. Gramlich, D. M. Hammond, T. Carell, *Org. Lett.* **2006**, 8, 3639-3642. *Click Chemistry as a Reliable Method for the High-Density Postsynthetic Functionalization of Alkyne-Modified DNA.*
- [92] C. T. Wirges, P. M. E. Gramlich, K. Gutsmedl, J. Gierlich, G. A. Burley, T. Carell, *QSAR Comb. Sci.* **2007**, 26, 1159-1164. *Pronounced Effect of DNA Hybridization on Click Reaction Efficiency.*
- [93] J. Gierlich, K. Gutsmedl, P. M. E. Gramlich, A. Schmidt, G. A. Burley, T. Carell, *Chem. Eur. J.* **2007**, 13, 9486-9494. *Synthesis of Highly Modified DNA by a Combination of Pcr with Alkyne-Bearing Triphosphates and Click Chemistry.*
- [94] K. Brunner, J. Harder, T. Halbach, J. Willibald, F. Spada, F. Gnerlich, K. Sparrer, A. Beil, L. Möckl, C. Bräuchle, *et al.*, *Angew. Chem. Int. Ed.* **2015**, 54, 1946-1949. *Cell-Penetrating and Neurotargeting Dendritic Sirna Nanostructures.*
- [95] T. Ami, K. Fujimoto, *ChemBioChem* **2008**, 9, 2071-2074. *Click Chemistry as an Efficient Method for Preparing a Sensitive DNA Probe for Photochemical Ligation.*
- [96] G. A. Burley, J. Gierlich, M. R. Mofid, H. Nir, S. Tal, Y. Eichen, T. Carell, *J. Am. Chem. Soc.* **2006**, 128, 1398-1399. *Directed DNA Metallization.*
- [97] R. Kumar, A. El-Sagheer, J. Tumpene, P. Lincoln, L. M. Wilhelmsson, T. Brown, *J. Am. Chem. Soc.* **2007**, 129, 6859-6864. *Template-Directed Oligonucleotide Strand Ligation, Covalent Intramolecular DNA Circularization and Catenation Using Click Chemistry.*
- [98] H. Krishna, M. H. Caruthers, *J. Am. Chem. Soc.* **2012**, 134, 11618-11631. *Alkynyl Phosphonate DNA: A Versatile “Click” Able Backbone for DNA-Based Biological Applications.*
- [99] A. H. El-Sagheer, R. Kumar, S. Findlow, J. M. Werner, A. N. Lane, T. Brown, *ChemBioChem* **2008**, 9, 50-52. *A Very Stable Cyclic DNA Miniduplex with Just Two Base Pairs.*
- [100] A. H. El-Sagheer, T. Brown, *J. Am. Chem. Soc.* **2009**, 131, 3958-3964. *Synthesis and Polymerase Chain Reaction Amplification of DNA Strands Containing an Unnatural Triazole Linkage.*
- [101] C. N. Birts, A. P. Sanzone, A. H. El-Sagheer, J. P. Blaydes, T. Brown, A. Tavassoli, *Angew. Chem. Int. Ed.* **2014**, 53, 2362-2365. *Transcription of Click-Linked DNA in Human Cells.*
- [102] V. I. Claessen, H. Engelkamp, Christianen, Peter C.M., J. C. Maan, R. J. M. Nolte, K. Blank, A. E. Rowan, *Annu. Rev. Anal. Chem.* **2010**, 3, 319-340. *Single-Biomolecule Kinetics: The Art of Studying a Single Enzyme.*
- [103] M. L. Metzker, *Nat. Rev. Genet.* **2010**, 11, 31-46. *Sequencing Technologies [Mdash] the Next Generation.*
- [104] J. S. Bartlett, in *Molecular Diagnosis of Cancer*, Vol. 97 (Eds.: J. Roulston, J. S. Bartlett), Humana Press, **2004**, pp. 77-87.
- [105] E. Fischer, B. Helferich, *Ber. Dtsch. Chem. Ges.* **1914**, 47, 210-235. *Synthetische Glucoside Der Purine.*
- [106] H. J. Schaeffer, S. Gurwara, R. Vince, S. Bittner, *J. Med. Chem.* **1971**, 14, 367-369. *Novel Substrate of Adenosine Deaminase.*
- [107] H. J. Schaeffer, L. Beauchamp, P. de Miranda, G. B. Elion, D. J. Bauer, P. Collins, *Nature* **1978**, 272, 583-585. *9-(2-Hydroxyethoxymethyl)Guanine Activity against Viruses of the Herpes Group.*

- [108] M. R. Boyd, T. H. Bacon, D. Sutton, M. Cole, *Antimicrob. Agents Chemother.* **1987**, *31*, 1238-1242. *Antiherpesvirus Activity of 9-(4-Hydroxy-3-Hydroxy-Methylbut-1-Yl)Guanine (Brl 39123) in Cell Culture.*
- [109] G. B. Elion, P. A. Furman, J. A. Fyfe, P. De Miranda, L. Beauchamp, H. J. Schaeffer, *Proc. Natl. Acad. Sci. USA* **1977**, *74*, 5716-5720. *Selectivity of Action of an Antiherpetic Agent, 9-(2-Hydroxyethoxymethyl) Guanine.*
- [110] J. Balzarini, A. Karlsson, S. Aquaro, C. F. Perno, D. Cahard, L. Naesens, E. De Clercq, C. McGuigan, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 7295-7299. *Mechanism of Anti-Hiv Action of Masked Alaninyl D4t-Mp Derivatives.*
- [111] E. Murakami, T. Tolstykh, H. Bao, C. Niu, H. M. M. Steuer, D. Bao, W. Chang, C. Espiritu, S. Bansal, A. M. Lam, *et al.*, *J. Biol. Chem.* **2010**, *285*, 34337-34347. *Mechanism of Activation of Psi-7851 and Its Diastereoisomer Psi-7977.*
- [112] T. K. Warren, J. Wells, R. G. Panchal, K. S. Stuthman, N. L. Garza, S. A. Van Tongeren, L. Dong, C. J. Retterer, B. P. Eaton, G. Pegoraro, *et al.*, *Nature* **2014**, *508*, 402-405. *Protection against Filovirus Diseases by a Novel Broad-Spectrum Nucleoside Analogue Bcx4430.*
- [113] F. H. C. Crick, L. Barnett, S. Brenner, R. J. Watts-Tobin, *Nature* **1961**, *192*, 1227-1232. *General Nature of the Genetic Code for Proteins.*
- [114] A. Rich, (Eds.: M. Kasha, B. Pullman), Academic Press NY, **1962**, pp. 103-126.
- [115] H. P. Rappaport, *Nucleic Acids Res.* **1988**, *16*, 7253-7267. *The 6-Thioguanine/5-Methyl-2-Pyrimidinone Base Pair.*
- [116] C. Switzer, S. E. Moroney, S. A. Benner, *J. Am. Chem. Soc.* **1989**, *111*, 8322-8323. *Enzymatic Incorporation of a New Base Pair into DNA and Rna.*
- [117] H. P. Rappaport, *Biochemistry* **1993**, *32*, 3047-3057. *Replication of the Base Pair 6-Thioguanine/5-Methyl-2-Pyrimidinone with the Large Klenow Fragment of Escherichia Coli DNA Polymerase I.*
- [118] J. A. Piccirilli, S. A. Benner, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* **1990**, *343*, 33-37. *Enzymatic Incorporation of a New Base Pair into DNA and Rna Extends the Genetic Alphabet.*
- [119] J. D. Bain, C. Switzer, R. Chamberlin, S. A. Benner, *Nature* **1992**, *356*, 537-539. *Ribosome-Mediated Incorporation of a Non-Standard Amino Acid into a Peptide through Expansion of the Genetic Code.*
- [120] C. Y. Switzer, S. E. Moroney, S. A. Benner, *Biochemistry* **1993**, *32*, 10489-10496. *Enzymic Recognition of the Base Pair between Isocytidine and Isoguanosine.*
- [121] A. M. Sismour, S. A. Benner, *Nucleic Acids Res.* **2005**, *33*, 5640-5646. *The Use of Thymidine Analogs to Improve the Replication of an Extra DNA Base Pair: A Synthetic Biological System.*
- [122] Z. Yang, F. Chen, J. B. Alvarado, S. A. Benner, *J. Am. Chem. Soc.* **2011**, *133*, 15105-15112. *Amplification, Mutation, and Sequencing of a Six-Letter Synthetic Genetic System.*
- [123] N. A. Leal, H.-J. Kim, S. Hoshika, M.-J. Kim, M. A. Carrigan, S. A. Benner, *ACS Synth. Biol.* **2015**, *4*, 407-413. *Transcription, Reverse Transcription, and Analysis of Rna Containing Artificial Genetic Components.*
- [124] M. Ishikawa, I. Hirao, S. Yokoyama, *Tetrahedron Lett.* **2000**, *41*, 3931-3934. *Synthesis of 3-(2-Deoxy-B-D-Ribofuranosyl)Pyridin-2-One and 2-Amino-6-(N,N-Dimethylamino)-9-(2-Deoxy-B-D-Ribofuranosyl)Purine Derivatives for an Unnatural Base Pair.*
- [125] T. Fujiwara, M. Kimoto, H. Sugiyama, I. Hirao, S. Yokoyama, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2221-2223. *Synthesis of 6-(2-Thienyl)Purine Nucleoside Derivatives That Form Unnatural Base Pairs with Pyridin-2-One Nucleosides.*
- [126] I. Hirao, T. Ohtsuki, T. Fujiwara, T. Mitsui, T. Yokogawa, T. Okuni, H. Nakayama, K. Takio, T. Yabuki, T. Kigawa, *et al.*, *Nat. Biotechnol.* **2002**, *20*, 177-182. *An Unnatural Base Pair for Incorporating Amino Acid Analogs into Proteins.*
- [127] J. M. Sturtevant, S. A. Rice, E. P. Geiduschek, *Farad. Discuss.* **1958**, *25*, 138-149. *The Stability of the Helical DNA Molecule in Solution.*

- [128] S. Moran, R. X. F. Ren, Rumney, E. T. Kool, *J. Am. Chem. Soc.* **1997**, *119*, 2056-2057. *Difluorotoluene, a Nonpolar Isostere for Thymine, Codes Specifically and Efficiently for Adenine in DNA Replication.*
- [129] S. Moran, R. X.-F. Ren, E. T. Kool, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10506-10511. *A Thymidine Triphosphate Shape Analog Lacking Watson–Crick Pairing Ability Is Replicated with High Sequence Selectivity.*
- [130] I. Hirao, M. Kimoto, T. Mitsui, T. Fujiwara, R. Kawai, A. Sato, Y. Harada, S. Yokoyama, *Nat. Methods* **2006**, *3*, 729-735. *An Unnatural Hydrophobic Base Pair System: Site-Specific Incorporation of Nucleotide Analogs into DNA and RNA.*
- [131] A. K. Ogawa, Y. Wu, D. L. McMinn, J. Liu, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **2000**, *122*, 3274-3287. *Efforts toward the Expansion of the Genetic Alphabet: Information Storage and Replication with Unnatural Hydrophobic Base Pairs.*
- [132] S. Matsuda, J. D. Fillo, A. A. Henry, P. Rai, S. J. Wilkens, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer, P. G. Schultz, G. Spraggon, *et al.*, *J. Am. Chem. Soc.* **2007**, *129*, 10466-10473. *Efforts toward Expansion of the Genetic Alphabet: Structure and Replication of Unnatural Base Pairs.*
- [133] A. A. Henry, A. G. Olsen, S. Matsuda, C. Yu, B. H. Geierstanger, F. E. Romesberg, *J. Am. Chem. Soc.* **2004**, *126*, 6923-6931. *Efforts to Expand the Genetic Alphabet: Identification of a Replicable Unnatural DNA Self-Pair.*
- [134] S. Matsuda, A. M. Leconte, F. E. Romesberg, *J. Am. Chem. Soc.* **2007**, *129*, 5551-5557. *Minor Groove Hydrogen Bonds and the Replication of Unnatural Base Pairs.*
- [135] Y. Kim, A. M. Leconte, Y. Hari, F. E. Romesberg, *Angew. Chem.* **2006**, *118*, 7973-7976. *Stability and Polymerase Recognition of Pyridine Nucleobase Analogues: Role of Minor-Groove H-Bond Acceptors.*
- [136] A. M. Leconte, G. T. Hwang, S. Matsuda, P. Capek, Y. Hari, F. E. Romesberg, *J. Am. Chem. Soc.* **2008**, *130*, 2336-2343. *Discovery, Characterization, and Optimization of an Unnatural Base Pair for Expansion of the Genetic Alphabet.*
- [137] G. T. Hwang, F. E. Romesberg, *J. Am. Chem. Soc.* **2008**, *130*, 14872-14882. *Unnatural Substrate Repertoire of  $\alpha$ ,  $\beta$ , and  $\chi$  Family DNA Polymerases.*
- [138] D. A. Malyshev, K. Dhama, H. T. Quach, T. Lavergne, P. Ordoukhanian, A. Torkamani, F. E. Romesberg, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 12005-12010. *Efficient and Sequence-Independent Replication of DNA Containing a Third Base Pair Establishes a Functional Six-Letter Genetic Alphabet.*
- [139] D. A. Malyshev, K. Dhama, T. Lavergne, T. Chen, N. Dai, J. M. Foster, I. R. Correa, F. E. Romesberg, *Nature* **2014**, *509*, 385-388. *A Semi-Synthetic Organism with an Expanded Genetic Alphabet.*
- [140] S. Katz, *J. Am. Chem. Soc.* **1952**, *74*, 2238-2245. *The Reversible Reaction of Sodium Thymonucleate and Mercuric Chloride.*
- [141] Y. Miyake, H. Togashi, M. Tashiro, H. Yamaguchi, S. Oda, M. Kudo, Y. Tanaka, Y. Kondo, R. Sawa, T. Fujimoto, *et al.*, *J. Am. Chem. Soc.* **2006**, *128*, 2172-2173. *Mercury(II)-Mediated Formation of Thymine–Hg(II)–Thymine Base Pairs in DNA Duplexes.*
- [142] Y. Tanaka, S. Oda, H. Yamaguchi, Y. Kondo, C. Kojima, A. Ono, *J. Am. Chem. Soc.* **2007**, *129*, 244-245.  *$^{15}\text{N}$ – $^{15}\text{N}$  J-Coupling across Hg(II): Direct Observation of Hg(II)-Mediated T–T Base Pairs in a DNA Duplex.*
- [143] A. Ono, S. Cao, H. Togashi, M. Tashiro, T. Fujimoto, T. Machinami, S. Oda, Y. Miyake, I. Okamoto, Y. Tanaka, *Chem. Commun.* **2008**, 4825-4827. *Specific Interactions between Silver(I) Ions and Cytosine–Cytosine Pairs in DNA Duplexes.*
- [144] H. Urata, E. Yamaguchi, Y. Nakamura, S.-i. Wada, *Chem. Commun.* **2011**, *47*, 941-943. *Pyrimidine–Pyrimidine Base Pairs Stabilized by Silver(I) Ions.*
- [145] T. Funai, Y. Miyazaki, M. Aotani, E. Yamaguchi, O. Nakagawa, S.-i. Wada, H. Torigoe, A. Ono, H. Urata, *Angew. Chem.* **2012**, *124*, 6570-6572. *Ag(II) Ion Mediated Formation of a C–a Mismatch by DNA Polymerases.*

- [146] A. Ono, H. Togashi, *Angew. Chem. Int. Ed.* **2004**, *43*, 4300-4302. *Highly Selective Oligonucleotide-Based Sensor for Mercury(Ii) in Aqueous Solutions.*
- [147] R. Freeman, T. Finder, I. Willner, *Angew. Chem. Int. Ed.* **2009**, *48*, 7818-7821. *Multiplexed Analysis of Hg<sub>2</sub><sup>+</sup> and Ag<sup>+</sup> Ions by Nucleic Acid Functionalized Cdse/Zns Quantum Dots and Their Use for Logic Gate Operations.*
- [148] T. Funai, J. Nakamura, Y. Miyazaki, R. Kiriu, O. Nakagawa, S.-i. Wada, A. Ono, H. Urata, *Angew. Chem. Int. Ed.* **2014**, *53*, 6624-6627. *Regulated Incorporation of Two Different Metal Ions into Programmed Sites in a Duplex by DNA Polymerase Catalyzed Primer Extension.*
- [149] K. S. Park, C. Jung, H. G. Park, *Angew. Chem. Int. Ed.* **2010**, *49*, 9757-9760. *"Illusionary" Polymerase Activity Triggered by Metal Ions: Use for Molecular Logic-Gate Operations.*
- [150] E. Meggers, P. L. Holland, W. B. Tolman, F. E. Romesberg, P. G. Schultz, *J. Am. Chem. Soc.* **2000**, *122*, 10714-10715. *A Novel Copper-Mediated DNA Base Pair.*
- [151] K. Tanaka, M. Shionoya, *J. Org. Chem.* **1999**, *64*, 5002-5003. *Synthesis of a Novel Nucleoside for Alternative DNA Base Pairing through Metal Complexation.*
- [152] K. Tanaka, A. Tengeiji, T. Kato, N. Toyama, M. Shionoya, *Science* **2003**, *299*, 1212-1213. *A Discrete Self-Assembled Metal Array in Artificial DNA.*
- [153] Y. Takezawa, W. Maeda, K. Tanaka, M. Shionoya, *Angew. Chem. Int. Ed.* **2009**, *48*, 1081-1084. *Discrete Self-Assembly of Iron(III) Ions inside Triple-Stranded Artificial DNA.*
- [154] S. Johannsen, N. Megger, D. Böhme, K. O. Sigel, J. Müller, *Nat Chem* **2010**, *2*, 229-234. *Solution Structure of a DNA Double Helix with Consecutive Metal-Mediated Base Pairs.*
- [155] T. P. Yoon, E. N. Jacobsen, *Science* **2003**, *299*, 1691-1693. *Privileged Chiral Catalysts.*
- [156] G. H. Clever, Y. Söhl, H. Burks, W. Spahl, T. Carell, *Chemistry-A European Journal* **2006**, *12*, 8708-8718. *Metal-Salen-Base-Pair Complexes inside DNA: Complexation Overrides Sequence Information.*
- [157] G. H. Clever, T. Carell, *Angew. Chem. Int. Ed.* **2007**, *46*, 250-253. *Controlled Stacking of 10 Transition-Metal Ions inside a DNA Duplex.*
- [158] S. Liu, G. H. Clever, Y. Takezawa, M. Kaneko, K. Tanaka, X. Guo, M. Shionoya, *Angew. Chem.* **2011**, *123*, 9048-9052. *Direct Conductance Measurement of Individual Metallo-DNA Duplexes within Single-Molecule Break Junctions.*
- [159] S. L. Beaucage, R. P. Iyer, *Tetrahedron* **1992**, *48*, 2223-2311. *Advances in the Synthesis of Oligonucleotides by the Phosphoramidite Approach.*
- [160] R. Saiki, S. Scharf, F. Faloona, K. Mullis, G. Horn, H. Erlich, N. Arnheim, *Science* **1985**, *230*, 1350-1354. *Enzymatic Amplification of Beta-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia.*
- [161] W. P. C. Stemmer, *Nature* **1994**, *370*, 389-391. *Rapid Evolution of a Protein in Vitro by DNA Shuffling.*
- [162] A. K. Showalter, M.-D. Tsai, *Biochemistry* **2002**, *41*, 10571-10576. *A Reexamination of the Nucleotide Incorporation Fidelity of DNA Polymerases†.*
- [163] D. A. Malyshev, D. A. Pfaff, S. I. Ippoliti, G. T. Hwang, T. J. Dwyer, F. E. Romesberg, *Chem. Eur. J.* **2010**, *16*, 12650-12659. *Solution Structure, Mechanism of Replication, and Optimization of an Unnatural Base Pair.*
- [164] J. Bowers, J. Mitchell, E. Beer, P. R. Buzby, M. Causey, J. W. Efcavitch, M. Jarosz, E. Krzymanska-Olejnik, L. Kung, D. Lipson, et al., *Nat. Methods* **2009**, *6*, 593-595. *Virtual Terminator Nucleotides for Next-Generation DNA Sequencing.*
- [165] M. L. Metzker, *Nat. Biotechnol.* **2009**, *27*, 150-151. *Sequencing in Real Time.*
- [166] C. Falentin, D. Beaupère, G. Demailly, I. Stasik, *Tetrahedron* **2008**, *64*, 9989-9991. *New Approach to (-)-Polyoxamic Acid and 3,4-Diepipolyoxamic Acid from D-Lyxono-1,4-Lactone.*
- [167] S. V. Attwood, A. G. M. Barrett, *J. Chem. Soc., Perkin Trans. 1* **1984**, 1315-1322. *A Convenient Procedure for the Deoxygenation and Homologation of D-Ribose Derivatives.*
- [168] K. Y. K. C. Park, H. Tomiyasu, *Synthesis* **1999**, *12*, 2041-2044. *A High-Yield Synthesis of 4-Borono-DL-Phenylalanine.*



- [169] S. Pitsch, P. A. Weiss, L. Jenny, A. Stutz, X. Wu, *Helv. Chim. Acta* **2001**, 84, 3773-3795. *Reliable Chemical Synthesis of Oligoribonucleotides (Rna) with 2'-O-[(Triisopropylsilyl)Oxy]Methyl(2'-O-Tom)-Protected Phosphoramidites.*
- [170] C. R. Allerson, S. L. Chen, G. L. Verdine, *J. Am. Chem. Soc.* **1997**, 119, 7423-7433. *A Chemical Method for Site-Specific Modification of Rna: The Convertible Nucleoside Approach.*
- [171] F. Chavan, B. Madje, J. Bharad, M. Ubale, M. Ware, M. Shingare, N. Shinde, *Bull. Catal. Soc. Ind.* **2008**, 7, 41-45. *Silicagel Supported Nahso4 Catalyzed Organic Reaction: An Efficient Synthesis of Coumarins.*
- [172] R. Gopinath, S. J. Haque, B. K. Patel, *J. Org. Chem.* **2002**, 67, 5842-5845. *Tetrabutylammonium Tribromide (Tbatb) as an Efficient Generator of Hbr for an Efficient Chemoselective Reagent for Acetalization of Carbonyl Compounds.*
- [173] M. Münzel, U. Lischke, D. Stathis, T. Pfaffeneder, F. A. Gnerlich, C. A. Deiml, S. C. Koch, K. Karaghiosoff, T. Carell, *Chem. Eur. J.* **2011**, 17, 13782-13788. *Improved Synthesis and Mutagenicity of Oligonucleotides Containing 5-Hydroxymethylcytosine, 5-Formylcytosine and 5-Carboxylcytosine.*
- [174] A. Okamoto, K. Tainaka, I. Saito, *Tetrahedron Lett.* **2002**, 43, 4581-4583. *A Facile Incorporation of the Aldehyde Function into DNA: 3-Formylindole Nucleoside as an Aldehyde-Containing Universal Nucleoside.*
- [175] N. D. Abeydeera, C. S. Chow, *Biorg. Med. Chem.* **2009**, 17, 5887-5893. *Synthesis and Characterization of Modified Nucleotides in the 970 Hairpin Loop of Escherichia Coli 16s Ribosomal Rna.*
- [176] D. Graber, H. Moroder, R. Micura, *J. Am. Chem. Soc.* **2008**, 130, 17230-17231. *19f Nmr Spectroscopy for the Analysis of Rna Secondary Structure Populations.*
- [177] A. Somoza, J. Chelliserrykattil, E. T. Kool, *Angew. Chem.* **2006**, 118, 5116-5119. *The Roles of Hydrogen Bonding and Sterics in Rna Interference.*
- [178] A. C. Bajji, D. R. Davis, *Org. Lett.* **2000**, 2, 3865-3868. *Synthesis and Biophysical Characterization of Trnals,3 Anticodon Stem-Loop Rnas Containing the Mcm5s2u Nucleoside.*
- [179] T. P. Blaisdell, S. Lee, P. Kasaplar, X. Sun, K. L. Tan, *Org. Lett.* **2013**, 15, 4710-4713. *Practical Silyl Protection of Ribonucleosides.*
- [180] C. T. Martin, D. K. Muller, J. E. Coleman, *Biochemistry* **1988**, 27, 3966-3974. *Processivity in Early Stages of Transcription by T7 Rna Polymerase.*
- [181] Y. J. Seo, D. A. Malyshev, T. Lavergne, P. Ordoukhanian, F. E. Romesberg, *J. Am. Chem. Soc.* **2011**, 133, 19878-19888. *Site-Specific Labeling of DNA and RNA Using an Efficiently Replicated and Transcribed Class of Unnatural Base Pairs.*
- [182] C. D. Kaul, Ludwig-Maximilians-Universität (München), **2011**. *Replikation Und Amplifizierung Des Metall-Salen-Basenpaars.*
- [183] N. C. Horton, B. C. Finzel, *J. Mol. Biol.* **1996**, 264, 521-533. *The Structure of an RNA/DNA Hybrid: A Substrate of the Ribonuclease Activity of Hiv-1 Reverse Transcriptase.*
- [184] K. Schneider, B. T. Chait, *Nucleic Acids Res.* **1995**, 23, 1570-1575. *Increased Stability of Nucleic Acids Containing 7-Deaza-Guanosine and 7-Deaza-Adenosine May Enable Rapid DNA Sequencing by Matrix-Assisted Laser Desorption Mass Spectrometry.*
- [185] M. Frugier, C. Florentz, M. W. Hosseini, J.-M. Lehn, R. Giegé, *Nucleic Acids Res.* **1994**, 22, 2784-2790. *Synthetic Polyamines Stimulate in Vitro Transcription by T7 Rna Polymerase.*
- [186] R. Sousa, Y. J. Chung, J. P. Rose, B.-C. Wang, *Nature* **1993**, 364, 593-599. *Crystal Structure of Bacteriophage T7 Rna Polymerase at 3.3 [Angst] Resolution.*
- [187] J. Huang, L. G. Briebe, R. Sousa, *Biochemistry* **2000**, 39, 11571-11580. *Misincorporation by Wild-Type and Mutant T7 Rna Polymerases: Identification of Interactions That Reduce Misincorporation Rates by Stabilizing the Catalytically Incompetent Open Conformation.*
- [188] N. Joubert, M. Urban, R. Pohl, M. Hocek, *Synthesis* **2008**, 1918-1932. *Modular Synthesis of 4-Aryl- and 4-Amino-Substituted Benzene C-2'-Deoxyribonucleosides.*

- [189] A. Mielke, T. Roubíček, *Multiscale Model. Sim.* **2003**, *1*, 571-597. *A Rate-Independent Model for Inelastic Behavior of Shape-Memory Alloys.*
- [190] M. G. Donahue, J. N. Johnston, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5602-5604. *Preparation of a Protected Phosphoramidon Precursor Via an H-Phosphonate Coupling Strategy.*
- [191] D. L. McMinn, A. K. Ogawa, Y. Wu, J. Liu, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **1999**, *121*, 11585-11586. *Efforts toward Expansion of the Genetic Alphabet: DNA Polymerase Recognition of a Highly Stable, Self-Pairing Hydrophobic Base.*
- [192] S. Matsuda, A. A. Henry, F. E. Romesberg, *J. Am. Chem. Soc.* **2006**, *128*, 6369-6375. *Optimization of Unnatural Base Pair Packing for Polymerase Recognition.*
- [193] L. A. Haff, I. P. Smirnov, *Genome Res.* **1997**, *7*, 378-388. *Single-Nucleotide Polymorphism Identification Assays Using a Thermostable DNA Polymerase and Delayed Extraction Maldi-Tof Mass Spectrometry.*
- [194] A.-C. Syvänen, K. Aalto-Setälä, L. Harju, K. Kontula, H. Söderlund, *Genomics* **1990**, *8*, 684-692. *A Primer-Guided Nucleotide Incorporation Assay in the Genotyping of Apolipoprotein E.*
- [195] C. Yu, A. A. Henry, F. E. Romesberg, P. G. Schultz, *Angew. Chem.* **2002**, *114*, 3997-4000. *Polymerase Recognition of Unnatural Base Pairs.*
- [196] I. Hirao, M. Kimoto, T. Mitsui, T. Fujiwara, R. Kawai, A. Sato, Y. Harada, S. Yokoyama, *Nucleic Acids Symp. Ser.* **2006**, *50*, 33-34. *An Unnatural Base Pair System for in Vitro Replication and Transcription.*
- [197] K. Betz, D. A. Malyshev, T. Lavergne, W. Welte, K. Diederichs, T. J. Dwyer, P. Ordoukhanian, F. E. Romesberg, A. Marx, *Nat. Chem. Biol.* **2012**, *8*, 612-614. *Klentaq Polymerase Replicates Unnatural Base Pairs by Inducing a Watson-Crick Geometry.*
- [198] J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, *Nature* **1998**, *391*, 304-307. *Visualizing DNA Replication in a Catalytically Active Bacillus DNA Polymerase Crystal.*
- [199] A. G. Steinig, M. J. Mulvihill, J. Wang, D. S. Werner, Q. Weng, J. Kan, H. Coate, X. Chen, *U.S. Patent No 20090197862 A1*, **2009**, p. 32.
- [200] F. Durola, D. Hanss, P. Roesel, J.-P. Sauvage, O. S. Wenger, *Eur. J. Org. Chem.* **2007**, *2007*, 125-135. *A New Family of Biisoquinoline Chelates.*
- [201] M. A. Lemmon, J. Schlessinger, *Cell* **2010**, *141*, 1117-1134. *Cell Signaling by Receptor-Tyrosine Kinases.*
- [202] B. M. Babior, *Blood* **1999**, *93*, 1464-1476. *Nadph Oxidase: An Update.*
- [203] P. C. van Weeren, K. M. T. de Bruyn, A. M. M. de Vries-Smits, J. van Lint, B. M. T. Burgering, *J. Biol. Chem.* **1998**, *273*, 13150-13156. *Essential Role for Protein Kinase B (Pkb) in Insulin-Induced Glycogen Synthase Kinase 3 Inactivation: Characterization of Dominant-Negative Mutant of Pkb.*
- [204] L. N. Johnson, D. Barford, *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 199-232. *The Effects of Phosphorylation on the Structure and Function of Proteins.*
- [205] E. Paredes, S. R. Das, *ChemBioChem* **2011**, *12*, 125-131. *Click Chemistry for Rapid Labeling and Ligation of Rna.*
- [206] D. Schulz, J. M. Holstein, A. Rentmeister, *Angew. Chem. Int. Ed.* **2013**, *52*, 7874-7878. *A Chemo-Enzymatic Approach for Site-Specific Modification of the Rna Cap.*
- [207] L. D. Lavis, R. T. Raines, *ACS Chem. Biol.* **2008**, *3*, 142-155. *Bright Ideas for Chemical Biology.*
- [208] T. A. Clark, I. A. Murray, R. D. Morgan, A. O. Kislyuk, K. E. Spittle, M. Boitano, A. Fomenkov, R. J. Roberts, J. Korlach, *Nucleic Acids Res.* **2012**, *40*, e29-e29. *Characterization of DNA Methyltransferase Specificities Using Single-Molecule, Real-Time DNA Sequencing.*
- [209] S. Koren, M. C. Schatz, B. P. Walenz, J. Martin, J. Howard, G. Ganapathy, Z. Wang, D. A. Rasko, W. R. McCombie, E. D. Jarvis, *et al.*, *Nat. Biotechnol.* **2012**, *30*, 693-700. *Hybrid Error Correction and De Novo Assembly of Single-Molecule Sequencing Reads.*
- [210] C.-X. Song, T. A. Clark, X.-Y. Lu, A. Kislyuk, Q. Dai, S. W. Turner, C. He, J. Korlach, *Nat. Methods* **2011**, *9*, 75-77. *Sensitive and Specific Single-Molecule Sequencing of 5-Hydroxymethylcytosine.*

- [211] N. Hardt, S. M. Hacker, A. Marx, *Org. Biomol. Chem.* **2013**, *11*, 8298-8305. *Synthesis and Fluorescence Characteristics of Atp-Based Fret Probes.*
- [212] S. Serdjukow, F. Kink, B. Steigenberger, M. Tomas-Gamasa, T. Carell, *Chem. Commun.* **2014**, *50*, 1861-1863. *Synthesis of [Gamma]-Labeled Nucleoside 5[Prime or Minute]-Triphosphates Using Click Chemistry.*
- [213] M. Morr, V. Wray, *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 1394-1396. *New Cyclic Derivatives of 3'-Amino-3'-Deoxy-Adenosine-5'-Diphosphate, -Triphosphate, and -Methylenebis(Phosphonate).*
- [214] D. G. Knorre, V. A. Kurbatov, V. V. Samukov, *FEBS Lett.* **1976**, *70*, 105-108. *General Method for the Synthesis of Atp Gamma-Derivatives.*
- [215] F. M. Kink, Ludwig-Maximilians-Universität (München), **2012**. *Kupfer-Katalysierte Huisgen-Cycloadditionen Zur Darstellung Von Fluoreszenzmarkierten Nukleosidtriphosphaten.*
- [216] K. Sivakumar, F. Xie, B. M. Cash, S. Long, H. N. Barnhill, Q. Wang, *Org. Lett.* **2004**, *6*, 4603-4606. *A Fluorogenic 1,3-Dipolar Cycloaddition Reaction of 3-Azidocoumarins and Acetylenes†.*
- [217] S. M. Hacker, M. Mex, A. Marx, *J. Org. Chem.* **2012**, *77*, 10450-10454. *Synthesis and Stability of Phosphate Modified Atp Analogues.*
- [218] S. Ameta, J. Becker, A. Jaschke, *Org. Biomol. Chem.* **2014**, *12*, 4701-4707. *Rna-Peptide Conjugate Synthesis by Inverse-Electron Demand Diels-Alder Reaction.*
- [219] C. Napoli, C. Lemieux, R. Jorgensen, *Plant Cell* **1990**, *2*, 279-289. *Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in Trans.*
- [220] K. Gavrilov, W. M. Saltzman, *Yale J. Biol. Med.* **2012**, *85*, 187-200. *Therapeutic Sirna: Principles, Challenges, and Strategies.*
- [221] V. Cameron, D. Soltis, O. C. Uhlenbeck, *Nucleic Acids Res.* **1978**, *5*, 825-833. *Polynucleotide Kinase from a T4 Mutant Which Lacks the 3' Phosphatase Activity.*
- [222] H. E. Gottlieb, V. Kotlyar, A. Nudelman, *J. Org. Chem.* **1997**, *62*, 7512-7515. *Nmr Chemical Shifts of Common Laboratory Solvents as Trace Impurities.*
- [223] V. P. K. John D. Williams, Philip E. Morris, and Leroy B. Townsend, *Org. Synth.* **2005**, *82*, 88-92. *D-Ribonolactone and 2,3-Isopropylidene(D-Ribonolactone).*
- [224] Y. Cen, A. A. Sauve, *J. Org. Chem.* **2009**, *74*, 5779-5789. *Diastereocontrolled Electrophilic Fluorinations of 2-Deoxyribonolactone: Syntheses of All Corresponding 2-Deoxy-2-Fluorolactones and 2'-Deoxy-2'-Fluoro-Nad+S.*
- [225] U. K. Laemmli, *Nature* **1970**, *227*, 680-685. *Cleavage of Structural Proteins During the Assembly of the Head of Bacteriophage T4.*
- [226] W. Kabsch, *J. Appl. Crystallogr.* **1993**, *26*, 795-800. *Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants.*
- [227] A. J. McCoy, R. W. Grosse-Kunstleve, L. C. Storoni, R. J. Read, *Acta Crystallogr. Sect. D* **2005**, *61*, 458-464. *Likelihood-Enhanced Fast Translation Functions.*
- [228] P. Emsley, K. Cowtan, *Acta Crystallogr. Sect. D* **2004**, *60*, 2126-2132. *Coot: Model-Building Tools for Molecular Graphics.*
- [229] N. Collaborative Computational Project, *Acta Crystallogr. D Biol. Crystallogr.* **1994**, *50*, 760-763. *The Ccp4 Suite: Programs for Protein Crystallography.*
- [230] G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr. Sect. D* **1997**, *53*, 240-255. *Refinement of Macromolecular Structures by the Maximum-Likelihood Method.*